National Defence / Defense nationale

DEFENCE AND CIVIL INSTITUTE OF ENVIRONMENTAL MEDICINE

Canada

December 1987 DCIEM NO. 87-RR-57

A PRELIMINARY EXAMINATION OF MENTAL WORKLOAD, ITS MEASUREMENT AND PREDICTION

P. S. Eisen

K. C. Hendy

Defence and Civil Institute of Environmental Medicine
1133 Shepard Avenue West
P.O. Box 2000
Downsview, Ontario
M3M 3B9

DEPARTMENT OF NATIONAL DEFENCE - CANADA

ABSTRACT

A principal focus of DCIEM's proposed work in human operator workload is the area of subjective mental load. This Report details a literature review, reviews the objectives of mental workload research, and defines terms which are related to the study of mental workload. A conceptual model of human cognitive behaviour is advanced which draws both from multiple resource theory and from hierarchical taxonomies of skilled behaviour. Potential indicators of mental workload are discussed and a set of criteria for judging workload metrics is described. Following this, a summary of available workload measures is presented. Finally, the report includes a discussion and recommendations for further research which is directed towards the validation of workload metrics in applied field settings and the establishment of the limitations of various measurement techniques.

Table of Contents

1. INTRODUCTION 1
2. OBJECTIVES OF MENTAL WORKLOAD MEASUREMENT 1
3. INDICANTS OF MENTAL WORKLOAD 2
4. MODELS FOR MENTAL WORKLOAD 5
4.1. Wickens' Multiple Resources Model 5
4.2. Rasmussen's Taxonomy of Human Cognitive Behaviour 7
4.3. A Modified Multiple Resources Model 8
5. WORKLOAD METRIC SELECTION CRITERIA 10
6. MEASURES OF MENTAL WORKLOAD 12
6.1. Behavioural Measures 12
6.1.1. Primary task performance 12
6.1.2. Spare mental capacity 14
6.1.3. Subjective Opinions 15
6.2. Analytical Procedures 17
6.3. Physiological Correlates 18
7. SUBJECTIVE WORKLOAD METRICS 19
7.1. Subjective Workload Assessment Technique (SWAT) 19
7.1.1. Dimensions of SWAT 21
7.1.2. SWAT Scale development 21
7.1.3. Event scoring 22
7.1.4. Applicability of SWAT 22
7.1.5. Validity of SWAT 23
7.1.6. Projective SWAT 23
7.1.7. Refective SWAT 24
7.1.8. Further validation of SWAT 24
7.2. NASA Task Load Index (NASA-TLX) 25
7.3. Cooper-Harper Scale 26
7.4. Magnitude Estimation 26
8. DISCUSSION 29
9. CONCLUSIONS 31
Appendix 1: DEFINITION OF WORKLOAD RELATED TERMS 39

1. INTRODUCTION

The Systems Study Group in the Human Engineering Section at DCIEM has initiated a programme of research (Hendy 1986) which has, as a major focal point, a concern with the evaluation and prediction of mental workload. This concern extends from the prediction of mental workload from mathematical models or systems simulations done early during the concept development stages of a project, through to the assessment of mental workload in prototype or operational systems. Of particular concern is the possibility that the various measures of workload which might be used at different points during development, test and evaluation might lead to different rankings for candidate systems.

As a starting point for the investigation of this topic, a literature survey was conducted. A prime objective in conducting this survey was the identification of a conceptual model for mental workload which would provide a common framework for two major classes of workload metric, *viz.*: (1) predictions derived from analytical procedures and mathematical simulation; and (2) subjective estimates of mental workload.

This Report presents the conclusions drawn from the literature survey and advances a model of mental workload which draws both on the popular multiple resources model of Wickens (1980) as well as Rasmussen's (1983) taxonomy of human behaviour.

The large amount of literature dealing with mental load has supplied a multiplicity of definitions for each of the terms used for describing and characterising its nature. Due to ambiguities of meaning, working definitions for these terms are established from the outset. These definitions are presented in Appendix 1.

2. OBJECTIVES OF MENTAL WORKLOAD MEASUREMENT

Many researchers in the field of workload measurement preface their reports with brief explanations as to the purpose of measuring mental load. They generally include concepts such as those expressed in the following two passages, taken from Helm (1981) and Williges and Wierwille (1979) respectively.

> 'Continued advancements in aircraft design and integrated systems avionics have significantly altered the role of pilot from that of skilled control operator to one of a manager of a complex system. Because of this change, emphasis on psychomotor control has been augmented by an interest in cognitive skills such as perception, memory, information processing, and decision making.' (p. 518)

> 'During the design, development, and test and evaluation of any advanced aircraft system...Care must be taken that the new system does not place unreasonable demands on the aircrew members by overwhelming them with too much information and too little time to process that information. Such considerations are often characterised as assessing the mental workload of the system operator.' (p. 549)

Gartner and Murphy (1976) list four main objectives of research into workload and fatigue. These include the comparison of alternative system or component designs with respect to demands imposed on the pilot and crew, and the establishment of functional relationships between task demands and effort required or the effectiveness of various effort investment strategies. They further provide a list of undesirable effects resulting from sustained pilot effort, for example, decrements in motivation, skill, and performance of the operator, along with a heightened level of stress. All of these factors act in concert to reduce the overall system performance. Gartner and Murphy give suggestions for an effective workload and fatigue management program, which

consists of generalities that are often difficult to implement. Key concepts mentioned, however, are the identification of symptoms and causes of workload in the cockpit, and the continued involvement of key personnel in workload management programmes.

Conversely, one may assess mental workload with the assumption that a lightly loaded operator will become less alert with time, thus reducing the level of performance. Here, the objective would be to raise the mental load to a level where the operator is busy enough to remain acute. This is common in tasks whose principal components require vigilance, such as monitoring flight instruments. Normally in aviation, however, the operator is required to timeshare several tasks and ***underloading*** is not a worry.

Moray (1982) states that '...a reliable way to measure ... workload is needed for reasons of safety, efficiency, wage-setting, and health' (p. 25). O'Donnell and Eggemeier (1982) suggest that the evaluation of alternative configurations of system components, with respect to cognitive demands, could be thoroughly carried out if a measure for mental load existed. In a separate paper, Eggemeier (1984) states that determining whether an overload can or does exist is the basic objective of all workload assessment. This approach neglects the option of measuring mental load below overload threshold values in order to set the imposed load on an operator at a value which optimises performance.

In summary, workload assessment techniques are employed in systems design and evaluation to attempt to provide answers to one or more of the following questions:

(1) Does the operator possess the capacities to perform the required tasks?

(2) Does the operator have enough spare capacity to take on additional tasks? If so, what tasks can be carried out, and to what level of performance?

(3) Does the operator have enough spare capacity to cope with emergency situations or other contingencies?

(4) Can the task or equipment be altered in any way so as to increase the amount of spare capacity of the operator?

(5) Can the task or equipment be modified in any way so as to increase the mental workload of the operator?

(6) How does the imposed workload of the new system compare with that of the old system?

3. INDICANTS OF MENTAL WORKLOAD

The concept of *mental workload*, or simply *mental load* was introduced in the performance literature to attempt to quantify and predict the human operator's performance level and overall system effectiveness in complex, dual, or multiple task situations. There is still no consensus on the definition of mental load (Moray 1982, O'Donnell and Eggemeier 1982, Williges and Wierwille 1979). A physiological approach, for example, explains mental load as the phenomenon that leads to the '...activation of an organism' (Haider 1972, p. 18). An equally broad definition is offered by Gerathewohl (1975), who states, 'One can...define workload as the sum of the task demands which can be clearly specified, plus the operator's response (and effort) to satisfy these demands' (p. C1-1), while a task analytic approach declares that '...workload is the ratio of the required performance time to the time available within the time constraints regulated by the mission' (Wingert 1973, p. 373).

It is, however, commonly accepted in the literature today that mental load is a multidimensional construct, although the factors which make up workload are yet to be determined. There have been many suggestions as to the dimensions of mental load, but almost all are based on pure conjecture with little or no substantive evidence behind the formulation of these theories. One

exception to this is the research based on the multiple resources theory of Wickens (see, for example, Derrick 1981 or Derrick and Wickens 1984), which hypothesises that 'The underlying structure of workload perceptions can be explained by task resource cost when resources are defined by stages, codes, and modalities.' (Derrick and Wickens 1984, p. 18). Another paper, by Skipper, Rieger and Wierwille (1986), showed that subjects rate workload on the basis of concepts that researchers generally consider to be associated with operator load, e.g., time pressure, difficulty, assessed performance and timesharing.

Derrick and Wickens (1984) used multi-dimensional scaling (MDS) techniques along with cluster analysis in an attempt to uncover the underlying dimensions of subjective workload. The main problem noted with this approach is that '...the identification of that binding property [of elements in a cluster or dimension] must come from further analysis...' (p. 15). Generally the investigator must assert the binding property and thus the analysis is prone to investigator bias. If the clusters or dimensions can be even partially accounted for by a particular theory (e.g., the multiple resources theory), the investigators may cease to look for alternatives. Derrick and Wickens used correlational analysis of subjective ratings and MDS of objective task parameters to derive the following three dimensions of mental load:

Dimension 1: Complex, resource-expensive versus simple, resource-inexpensive;

Dimension 2: Visually simple, good feedback versus auditorily simple, poor feedback; and

Dimension 3: Simple, time-free versus complex, time-committed.

The results are far from conclusive, although they do suggest that further research is warranted to better isolate the dimensions of subjective workload and to relate these dimensions to objective and subjective workload indicators.

Mental workload has been characterised as being a phenomenon of task input, operator internal state, response output, or any combination of these. As Rasmussen (1979) remarks, '...workload turns out to be a measure characterising a specific man-machine encounter, a situation, not the task itself.' (p. 33). Accordingly, many researchers (e.g., Aldrich, Craddock and McCracken 1984, Helm 1981, Johannsen 1979, O'Donnell and Eggemeier 1982) feel that a suitable measure of mental load must incorporate parameters from all of these domains.

Johannsen (1979) describes a scheme of operator load in which the segments are input load, operator effort, and performance. Various parameters for measuring each segment are mentioned. A physiological approach is taken by Rohmert (1971), who views the workload system as being made up of stressors, stresses, and strains. According to Rohmert, ***stressors*** are task inputs, ***stresses*** are the resultant effects of these inputs on the operator, and ***strains*** are the operator's responses, measured by a number of potential physiological indicators. It is the operator stresses which are most difficult to understand and estimate.

O'Donnell and Eggemeier (1982) present a framework for studying mental load that consists of task demands, operator variables, and response variables, along with corresponding indicators. They adopt the point of view that workload is '...viewed as a hypothetical construct which conveniently summarizes the multiple interactions which limit operator task performance. The workload of a mission cannot be thought of as independent of the mission, the system, or the particular individuals engaged in the operation.' (p. 4-1).

This approach to mental load is not inconsistent with Wickens' view of multiple resources (Derrick and Wickens 1984, Shingledecker 1984). Wickens' ***modality*** provides the input variables, with ***processing stages*** and ***codes*** providing both the operator centered variables and the output variables. Rephrased, the structure of studying mental workload in terms of inputs, operator-centred variables, and outputs, is embedded into the concept of Wickens' multiple resources theory.

The word ***indicator*** is defined by Klausner (1964) to be '...a term which gives an instruction about where to look for observable evidence about the concept referent.' Table 1 contains a list of verbal descriptors that have been suggested as possible indicators and/or parameters of mental load. The list is divided into three categories, which correspond to the principal components of the workload system — task demands, operator-centred variables, and performance criteria. Other than those marked, each of these potential indicators can be quantified by both objective and

subjective measures. Physiological indicators are not considered.

TABLE 1: INDICATORS OF MENTAL LOAD

Indicators of Mental Load		
Task Demands	Operator-Centred Variables	Performance Criteria
intensity of physical work	motivation †	probability of failure
information input rate *	effort	number of errors
time available	attitude towards task †	errors' influence on load
lead time constant *	personality †	risk
frequency of interruptions	general background †	precision required
interruptions' influence on load	aptitude	uncertainty of consequences
interference of auxiliary tasks	confusion †	conflict of demanded outcome
difficulty	anxiety †	difficulty
time stress	frustration †	criticality
stress	strain †	
quality of data	experience	
number of alternative solutions	attentiveness/attention	
feedback adequacy	need for feedback	
planning required	proximity to threshold (overload) †	
frequency of decisions	mental tension †	
complexity of decisions required	pre-task confidence †	
input complexity	recognition (of loading level) †	
response complexity	anticipation	
input coherence/continuity		
pressure of responsibility		
degree of human intervention/control		
* objective measures only † subjective measures only		

In reviewing several workload studies, Derrick and Wickens (1984) found that when one parameter of a task is varied, it is not uncommon for several workload indicators to covary. However, the manipulation of two or more parameters among workload conditions can cause dissociation among different indicators, especially when the parameters altered have offsetting effects on the actual mental load. This gives rise to the assertion that the various indicators reflect different parameters that may be associated with workload, but they do not individually constitute a measure of mental load. Further, a good measure of workload would likely be comprised of multiple indicators in order to account for the various parameters that constitute the workload system.

4. MODELS FOR MENTAL WORKLOAD

The successful application of scientific principles to the engineering design of human-machine systems generally requires models of a mathematical, iconic, or symbolic nature. Rouse (1980) describes four major uses of models:

(1) the model development procedure itself provides many insights into the properties of the system being modelled;

(2) the model is a concise tool for synthesising and analysing data;

(3) models are the basis of a systematic formation of hypotheses that can be experimentally tested, in order to resolve key issues regarding a system, and on which sensitivity analyses of various system parameters can be performed; and

(4) models are predictive instruments which allow for qualitative and quantitative measures to be considered early in the design stage of the human-machine system, minimizing the amount of post-implementation retrofitting that is carried out.

Shingledecker, Crabtree and Acton (1982) suggest that one of the '...major thrust areas which could contribute to the design of a coherent workload assessment technology [is] a model which would accurately characterise the functional operation of the human information processing system, and describe the nature of the capacities of this system which limit performance.' (p. 648). Further, Rasmussen (1979) asserts that mental workload '...cannot be evaluated unless a description or model of the operator's mental process is available.' (p. 30).

A modified multiple resources model is proposed which draws both from Wickens' (1980) multiple resource theory and on Rasmussen's (1984) taxonomy of human cognitive behaviour. The main objectives of this model are to assist in the conceptual understanding of the issues involved in mental workload assessment, and to provide a framework by which valid and reliable comparative evaluations of alternative workload measurement techniques can be made.

4.1. Wickens' Multiple Resources Model

Wickens' multiple resources model is often presented (Derrick and Wickens 1984) in the rather confusing form shown in Figure 1. The representation in Figure 1 attempts to display resource structures as defined by the categories of processing stages, processing modalities, and processing codes. These categories were inferred from several performance studies cited by Wickens (1980). Derrick (1981) and Derrick and Wickens (1984) hypothesised that the categorizations which apparently form the basis for resource structures may also be the underlying dimensions along which subjective estimates of workload lie.

A fundamental problem in interpretation derives from the incomplete division of Figure 1 into a cellular structure. It is not clear from the diagram whether 6, 7 or 8 cells are implied. Wickens does not wish to make a distinction between the resources used for perceptual (and it is assumed that this includes the *encoding* activity) and central-processing activities, therefore it would seem to be appropriate to continue the line which distinguishes between input modalities in Figure 1, through to meet the vertical line which separates the encoding/central processing cells from those which are associated with response selection and action. Once this is done, 6 cells are unequivocally defined. A potential problem with this solution is the implication that differentiated resources exist for auditory and visual processing beyond encoding.

The model is formed on the basic premise that resource competition causes interference, which leads to a degradation in performance in one or more of the tasks being carried out. It is important to qualify this, however, in light of the work done by Norman and Bobrow (1975) who distinguish between resource-limited and data-limited processes, *viz.*:

> 'Whenever an increase in the amount of processing resources can result in improved performance, we say that the task (or performance on that task) is *resource-limited*...Whenever performance is independent of processing resources, we say that the task is *data-limited*.' (p. 46)

This dichotomy is depicted in the performance-resource function (PRF) displayed as Figure 2. The PRF was introduced by Norman and Bobrow as an explanation for the interference observed in dual task performance when the tasks are assummed to be competing for resources. In the particular situation represented in Figure 2, a hypothetical task requires a minimum threshold value of

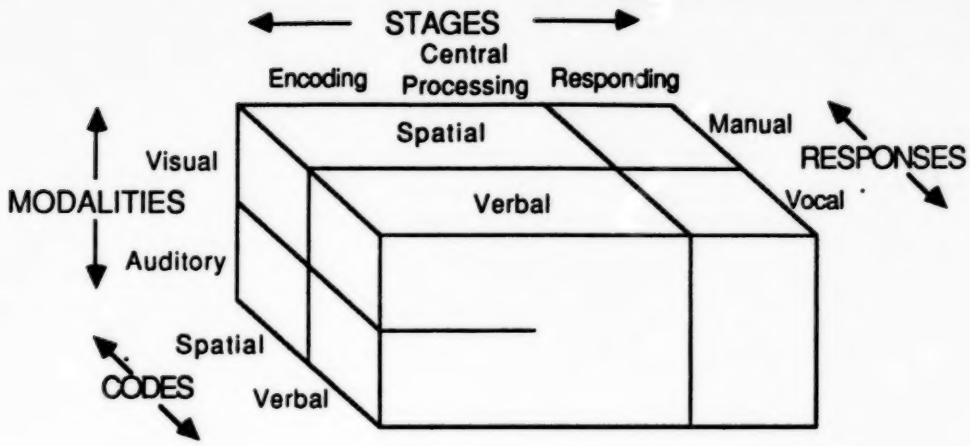

Figure 1. *Wickens' Multiple Resources Model*

resource allocation (R_{min}) in order to achieve any output whatsoever. From this point until a specific amount of resources are allocated to the task (at R_{dl}), performance is resource-limited, i.e. performance on the task is limited by how much processing capacity is devoted to the task. From R_{dl} until the limit of resources (R_l), the process is data-limited, i.e. the quality of the input signal or the quality of the stored representation in memory determines the maximum performance that can be achieved (Norman and Bobrow 1975). In the data limited region, performance can not be improved by the allocation of additional processing resources to the task.

The possibility that performance is data rather than resource limited, places constraints on the region in which the multiple resource model is useful as a predictive tool. The investigator has the additional problem of determining the region of the performance-resource curve in which a subject may be operating, for each of the tasks that are being performed. The extent of interference depends on whether the process is resource- or data-limited.

It is often assumed that '...vocal responses are usually verbal, and manual responses are normally spatially guided.' (Derrick and Wickens 1984, p. 5). There are many tasks, however, that cannot be adequately described under this assumption. Consider a situation where the operator is required to view various two-dimensional images and then record the shapes of these images by typing in an appropriate response on a keyboard. The input modality is visual and the response must be considered manual. However, it is not clear whether the coding would be designated as spatial or verbal. It is possible that the shape of the image is perceived and coded spatially, then translated through some form of cross-code transformer to verbal information, from which the operator can determine how to correctly spell the answer. This verbal information may get recoded spatially in order for the operator to execute the correct manual responses. Wickens' multiple resource model does not readily cope with code transformations such as these. Alternative coding schemes have been suggested in the literature (e.g., Harris and Haber 1963, Kahneman 1973, Wickens 1980, Wickens 1984) and there is evidence supporting many of the proposed theories. The ambiguities that remain unresolved in this area leave little confidence resting in a scheme that considers spatial and verbal representations and no others, as does Wickens' model.

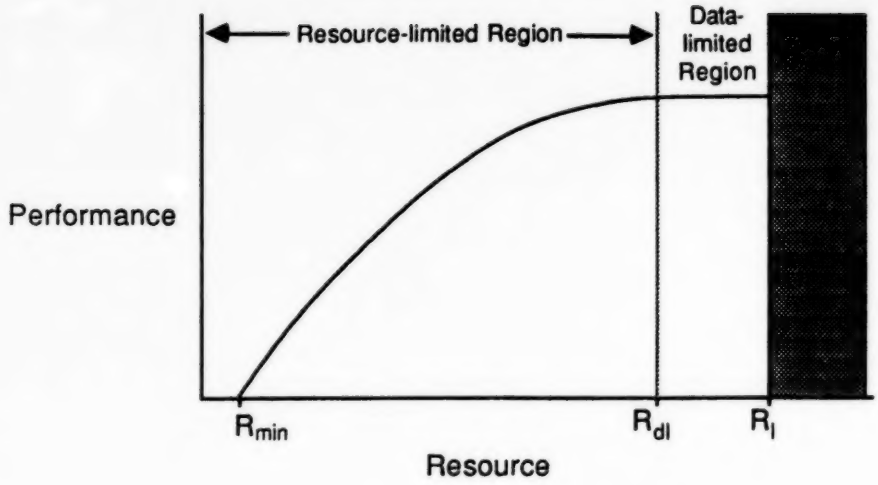

Figure 2. *The Performance-Resource Function*

In addition, the multiple resources model neglects the effects of stimulus-central processing-response (SCR) compatibility (Robinson and Eberts 1987). Performance in a dual-task situation with little resource competition (interference) may be degraded by poor SCR compatibility; performance in a high-interference dual-task scenario can be ameliorated by high SCR compatibility. Thus workload prediction in specific cases is a complex matter.

On the basis of these conclusions, the multiple resources model of Figure 1 is appears to be overly limiting in applicability to various task and workload situations. The utility of a multiple resources model in explaining human performance is well accepted (Navon and Gopher 1979), although its use in the predictive sense is yet to be established. Accordingly, a modified version of Wickens' model that maintains the concept of resource structures, but removes some constraints on their dimensional formation, is proposed in this report. This model is described in Section 4.3.

4.2. Rasmussen's Taxonomy of Human Cognitive Behaviour

Rasmussen (1983) presents a categorisation of cognitive behaviour that is a formidable conceptual aid in understanding human performance. Rasmussen does not claim this taxonomy to be a model, as it is not sufficiently descriptive. It is, however, well worth considering as a modifier for models that cannot account for various levels of information processing depth, or degrees of resource cost.

Rasmussen classifies cognitive behaviour into skill-based, rule-based, and knowledge-based behaviours, which can be thought of as representing levels of information processing depth. These levels take the form of layers which process information at increasing levels of abstraction, providing some commonality with the concept of layered protocols (Taylor 1986). A description of each level or layer follows.

Skill-Based Behaviour

According to Rasmussen (1983) 'The skill-based behaviour represents sensory-motor performance during acts or activities which, following a statement of intention, take place without conscious control as smooth, automated, and highly integrated patterns of information.' (p. 258). Skill-based behaviour is generally carried out in an open loop manner, i.e. any feedback that occurs is for updating mental models of the activity (learning) rather than for initiating corrective actions. A familiar example of a skilled behaviour is the signing of one's name, which is carried out automatically and without feedback by highly practised individuals. Skills constitute the lowest tier of the layered processing, in which the dominance of automation and reflex action imply little or no central processing.

Rule-Based Behaviour

Rule based behaviour is characterised by situation-action pairs based on pattern recognition. The human operator invokes a previously stored rule (either learned through experience or communicated from others) to respond consistently to recognizable situations. This behaviour forms the next layer of the processing, where the human operator is involved in pattern recognition, association and recollection of the appropriate response; which is then initiated, generally, as a skill-based behaviour. A common example of a rule-based action is the decision to stop after perceiving a stop sign while driving. The rule-based decision initiates the intention to stop, which is then performed at a skill-based level as a simple ballistic pedal movement. As the rule-based behaviour becomes more practised, responses tend towards integration and automation.

Knowledge Based Behaviour

This highest level of processing is also the most ambiguously defined. Here the operator is confronted with an unfamiliar situation which calls upon acquired skill and rule bases for decision making. Numerous alternatives may be considered and each tested against the goal until a suitable solution is achieved. Several processing layers make up this level of cognition. The ac ual layer or depth of processing depends on the complexity of the situation or problem, which accounts for the non-specific description of cognitive processes associated with knowledge-based behaviour.

The boundaries which separate these three levels of behaviour (which can also be termed ***automation***, ***association***, and ***cognition***) are fuzzy, and vary with the level of training and attention of the operator (Rasmussen 1983). This classification scheme seems to lend itself naturally toward ***post hoc*** analysis, but rates poorly as a predictor of cognitive behaviour. However, its utility as a conceptual tool towards understanding human behaviour is considerable and therfore it rates inclusion among the other concepts presented in this Section.

4.3. A Modified Multiple Resources Model

Likely the most effective structure to maintain in a model describing mental processes and workload is the three-component concept of task inputs, operator internal state, and response outputs (performance), as described previously. From a strictly operator-centred view, these components translate to perceptual input, central processing, and response outputs. Encoding and decoding functions are assumed to exist at the boundaries between these categories.

In this report the human operator is considered to be a multiple-resource parallel information processor. There are resource structures located within the perceptual input domain, and each sensory modality — visual, auditory, tactile, olfactory, gustatory, and kinaesthetic — draws on certain resource pools. It will be assumed that the resource pools are largely independent although the extent to which these reservoirs are differentiated is not known. Similarly, there are resource pools dedicated to response outputs. Each output medium — vocal, manual, eye movements and other motor functions — draws upon separate resources.

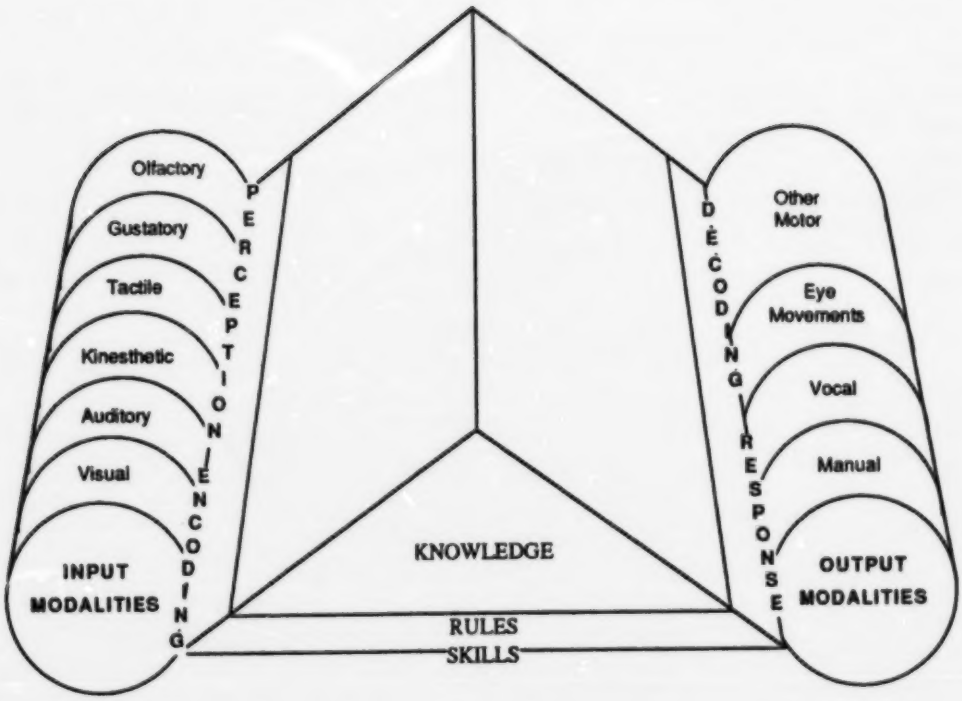

Figure 3. *A Modified Multiple Resources Model*

Central processing functions draw upon one or more resources as well. The specific nature of these resources is not known, however it is assumed that central resources are allocated on the basis of *processing level*. The depth or level of central processing is determined by the activity in which the operator is involved. Skill-based behaviours are unconscious automated acts that do not draw upon central processing resources to any significant degree. Rule-based actions require pattern matching and response-effecting resources and are carried out on a higher level of cognitive processing. Knowledge-based behaviour encompasses all levels of cognition above the layer of rules, including signal identification, decision-making, and planning. It is worthwhile to reiterate that this taxonomy is not strict , but is flexible to allow for peculiarities in operator strategies and behaviour.

It is assumed that there exists an executive resource whose function it is to allocate all other resources to demands. Whether this is a differentiated resource or resides in a common pool of central resources is unknown. A principal limiting function of cognitive performance is the ability to effectively allocate attentional resources.

A model which draws on the resources concept of Wickens but provides for a different allocation of resources is shown in Figure 3. The three component concept of task inputs, operator internal state and response outputs is retained but the resources associated with the stages of processing in Wickens' model are replaced by resources associated with the level of information processing.

In Figure 3 the input stage provides for separate resource pools associated with each of the sensory modalities. It is assumed that information is encoded and acted on by central processing

resources which are distinguished by their level of processing. The lowest level of processing, skill-based behaviour, is assumed to draw insignificantly upon resources, and therefore encompasses a negligible volume of the prism. Skill-based operations are seen as straightforward input-output activities. Rule-based behaviour requires minimal resources, portrayed by the thin slice of the prism allotted to this function. Finally, knowledge-based behaviour may require any resource in central processing, and at a relatively greater resource cost. If the operator chooses to respond, the response is effected, decoded, and performed in one or more of the possible output modalities listed.

This model suggests two fundamental dimensions of workload, *viz.*: a time component as suggested both by the serial nature (i.e. from input to output) of the model and by the possibility of resource competition; and an intensity or effort component which is related to the level of information processing required. It is suggested on the basis of this model that a *time* and *intensity* dimension should be present in any useful workload metric. Further, if different workload measurement or prediction techniques are used at various points in the system development programme, then these two fundamental aspects must be represented at all stages if the concept of workload is to stay constant.

5. WORKLOAD METRIC SELECTION CRITERIA

Many workload metrics have been proposed, dating back to the mid-1960's. Any researcher attempting to assess mental workload must choose one or more techniques according to the particular requirements of the investigation, and a sound knowledge of the advantages and limitations of each. A basis for judging potential metrics is presented in this Section.

Eggemeier (1984) recommends five criteria be used in selecting a workload metric: (1) sensitivity; (2) diagnosticity; (3) intrusiveness; (4) implementation requirements; and (5) operator acceptance. Gartner and Murphy (1976) propose a similar set of criteria under different verbal descriptors (but exclude implementation requirements), i.e.: (1) clarity of focus (distinguishing the phenomena of interest); (2) operational relevance (task demands and relationship with research); (3) practical significance; and (4) operator acceptance (including intrusiveness). The criteria chosen by Eggemeier are described below.

Sensitivity

A workload metric should possess sufficient sensitivity so as to be able to distinguish among several loading situations imposed on the operator by the system or system components. It has been stated that, 'The sensitivity criterion is ... an essential consideration in the choice of a workload measure, since the degree of sensitivity bears directly on the type of question that can be addressed by a technique.' (Eggemeier 1984, p. C7). A metric which can distinguish only between situations of very low workload and very high workload will be of little use in most applied settings.

Diagnosticity

Derrick and Wickens (1984) state that a '...measure is said to be diagnostic if it can specify the exact cause of the increased workload.' (p. 74). The criterion of diagnosticity finds its roots in the multiple resources explanation of human cognitive behaviour (Derrick and Wickens 1984, Navon and Gopher 1979, Sanders 1979, Shingledecker 1984), i.e. an appropriate workload assessment technique should be able to determine which of the resources are overloaded or near threshold in any loading situation. By successfully determining which resources are called upon, a measure could be used to predict which task or combination of tasks will induce an overload. These tasks

could then be examined more closely in an effort to reduce the imposed load.

The criteria of sensitivity and diagnosticity could be employed in tandem. A metric with high sensitivity could detect problematic situations; then a more diagnostic measure could be used to isolate the resources involved (Derrick and Wickens 1984, and Shingledecker 1984).

Intrusiveness

As a practical constraint, the technique chosen to assess mental load should in no way interfere with the performance of the primary task, nor with the actual load experienced by the operator (Gartner and Murphy 1976). In addition to the difficulties of interpretation which result when the phenomena of interest is disturbed, the safety of the operator may be compromised in certain task environments such as flight.

Implementation Requirements

Implementation requirements refer to the hardware, software, training, data analysis, and other equipment or procedures necessary to employ an assessment technique. Pragmatically, these factors should not be so cumbersome as to inhibit the successful application of a workload measure.

Operator Acceptance

As a final practical constraint, '...operator acceptance is important to ensure that an assessment technique will yield data that are representative of the load imposed by the task or system function in question. Assessment procedures that are perceived by operators as bothersome or artificial incur the risk of being ignored, [or] performed at substandard levels.' (Eggemeier 1984, p. C11). The effectiveness of the technique is thus compromised.

Reliability

In addition to the aforementioned considerations, the choice of a workload metric should take reliability into account. Here reliability refers to the ability of the measure to reproduce consistent results in several trials over an extended period of time.

Applicability

Two factors which relate to the applicability of a metric are:

(1) The ability of the measure to reproduce in the field the results obtained in the laboratory. A good human engineering method must possess this feature in order to provide useful results.

(2) The ability of the measure to produce valid results over a wide range of loading situations.

6. MEASURES OF MENTAL WORKLOAD

Scientists from a number of disciplines — experimental psychology, control engineering, mathematics, physiological psychology, and applied psychology — have all approached the problem of measuring mental workload from their own perspectives. In 1979 a conference was held in which leading researchers from these fields, attempted to establish a consensus on the definition and nature of mental load and on the most promising areas for further investigations. Along with

finding a surprising consensus among the disciplines regarding the value of various methods of measurement, the participants '...left the conference with a firm conviction that interdisciplinary co-operation is absolutely mandatory for progress.' (Moray 1979, p. vii).

Several reviews have facilitated the human engineer's task of surveying the literature in quest of an appropriate workload technique for a particular application (see, for example, AGARD-CPP-216 1977, AGARD-CPP-217 1977). Wierwille and Williges (1978) provide an excellent overview, in a user's guide format, of the available methods of measurement. Each method is briefly described, applications are mentioned and the limitations are outlined. The potential for further research of the technique is also evaluated. Finally, a step by step procedure for choosing a measurement method for any specific application is presented. The evaluations are based on the criteria discussed in Section 6, but include a more comprehensive list of practical constraints. Those not mentioned previously are:

- physical space required
- portability
- data transmission and recording
- experimental control
- integration into the human-machine system.

Meshkati, Hancock and Robertson's (1984) review of the mental load literature provides a thorough coverage of the work done up to 1980. Their review also presents information in the form of a users' guide. Other critical reviews have been produced by Eggemeier (1981) and Moray (1982).

The attempts to measure/predict mental load can be divided into three main categories, *viz.*: behavioural measures; analytical measures; and physiological correlates. The various approaches are discussed briefly below.

6.1. Behavioural Measures

Behavioural measures can be categorised into three groups: (1) primary task; (2) spare mental capacity; and (3) subjective opinions (see Figure 4). Each behavioural measure attempts to quantify one or more of the load indicators listed in Table 1. It would seem that, in order to obtain a workload value that is truly representative of the workload situation, at least one measure from each the categories *task demands*, *operator internal state* and *response output* should be considered.

6.1.1. Primary task performance

Primary task methods are designed to measure operator output (response) variables, or task performance. The underlying assumption is that the operator's performance is degraded by increasing mental workload (Wierwille and Williges 1978). However, as many researchers point out (e.g., Cooper and Harper 1969, Moray 1982), the human operator's ability to compensate with consistent system outputs for varied task demands and environmental conditions, often renders any performance measure useless as a workload indicator. In Singleton's (1971) words:

> 'Man is a complex set of interacting mechanisms with extensive feedback loops within and between systems. He must also be considered as an on-going total system continuously adjusting his patterns of internal and external activities so that the study of cause/effect relationships is essentially fallacious.' (p. 58)

The exception to this occurs in an overload situation, when the operator's performance is clearly degraded (Norman and Bobrow 1975). However, in most cases, Norman and Bobrow see this as being '...a smooth degradation on task performance rather than a calamitous failure.' (p. 45). In other words, the loading threshold level may not be readily measurable by primary task

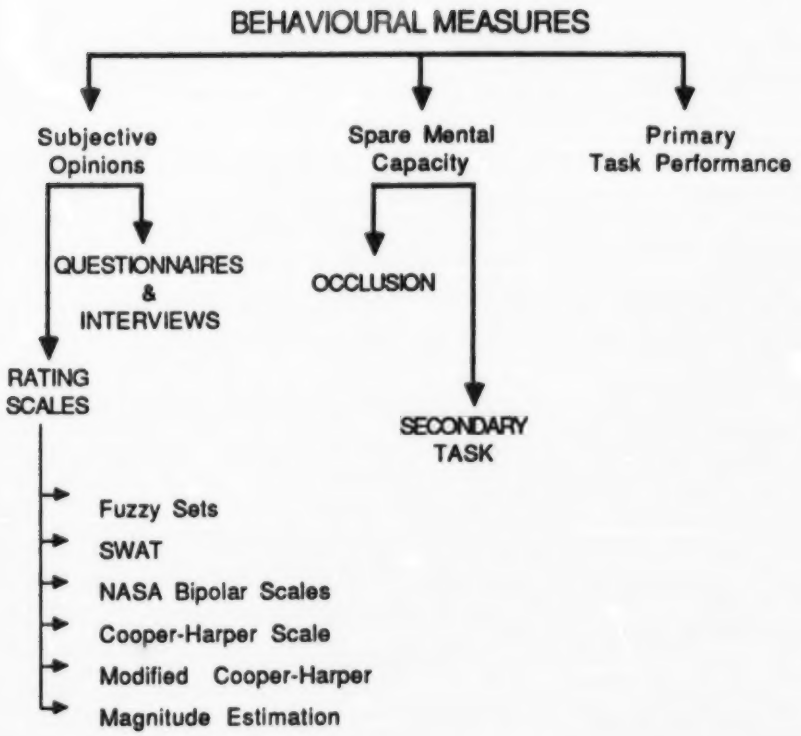

Figure 4. *Behavioural Measures of Mental Load*

techniques. Primary task measures are best suited to detecting the existence of an overload situation (Eggemeier 1984, Wierwille and Williges 1978); their merit is minimal when considering differences in workload below threshold.

A second shortcoming of primary task measures is that it is difficult to make valid comparisons among several indicators of task difficulty (Wickens 1984). For example, what is the change in control stability of a critical tracking task that produces a workload increase equivalent to that of increasing the information content of a choice reaction test from say 3 to 5 bits? This sort of comparison is useful when evaluating a multi-task environment, in search of workload tradeoffs to optimise the overall system performance.

6.1.2. Spare mental capacity

The largest body of knowledge generated thus far in workload assessment has been in the area of spare mental capacity (Wierwille and Williges 1978). Measures of spare mental capacity are based on the assumption that the human operator possesses a limited capacity in one or more resources which are called upon in the performance of tasks. These resources could reside in input, central or output processes; from encoding through to decoding and responding. If the task demands do not exceed the limits of the operator's abilities, the difference between the resources invoked and the total resources available can be considered to be spare mental resources, and are

therefore available for the performance of other cognitive activities. Thus, if the amount of spare mental resources can be determined while an operator is performing a task, the resources that are not available are assumed to be dedicated to the primary task. By this method, the cognitive demands of alternative primary tasks can be compared, and overload conditions (or conditions where the operator is at least fully loaded) can be detected. Methods to measure spare mental capacity that have been investigated thus far, can be categorised under two headings: (1) secondary tasks; or (2) occlusion (Wierwille and Williges 1978).

Secondary Tasks

According to Wierwille and Williges (1978), 'Most behavioral research approaches to estimating spare mental capacity have used secondary task procedures.' (p. 64). Detailed reviews of the secondary task literature are provided by Rolfe (1971) and Ogden, Levine and Eisner (1979). Ogden et al. provide a list of five types of decisions that can be made on the basis of secondary tasks:

(1) the ordering of difficulty of primary tasks;
(2) the evaluation of instrumentation and displays;
(3) the evaluation of operating conditions and procedures;
(4) the assessment of practice effects; and
(5) the selection of personnel.

To this list, diagnosis of resource allocation should be added. In favour of employing a battery of complementary assessment techniques to obtain a valid measure of mental load, Rolfe (1971) concludes his review of the field by declaring that '...the secondary task is no substitute for competent and comprehensive measurement of primary task performance. The technique should always be looked upon as a means of gathering additional information rather than an easy way of obtaining primary information.'

Occlusion

Occlusion is a technique for tasks in which vision is the primary information input modality. Often the operator's capacity for visual input is the most limiting function in the performance of a task, and demands exceeding this capacity will cause an overload (Senders 1983). Occlusion studies are based on the supposition that an operator performing a task is not continually viewing but is visually sampling the external world to pick up information that is required for the performance of the task. Control of the duration and frequency of sampling, by occluding the operator's view, can theoretically be used to determine the workload demands of the task. This is based on the premise that the more frequently or longer the operator needs to sample, the higher the workload or difficulty of the task (Wierwille and Williges 1978).

Proponents of the occlusion technique argue for its ability to accurately reflect attentional demand (in the sense of sampling rate rather than attentional effort), which in turn is an indicator of task difficulty. This seems plausible as long as vision is truly the foremost and limiting function of the operator's behaviour. However, should a significant amount of information processing be necessary for the successful completion of a task, occlusion measures would prove to be insensitive to actual workload.

Occlusion measures have the drawbacks of being both intrusive and potentially unsafe (Wierwille and Williges 1978). The operator may also adjust his strategy for performing the task, thereby changing the nature of the problem. As well, the technique cannot be applied to critical situations, where the occluded view might endanger the safety of the operator. A further limitation of the technique is its apparently poor sensitivity when compared with other workload measures (Hicks and Wierwille 1979).

6.1.3. Subjective Opinions

Subjective opinions are often employed in workload studies, either alone or in conjunction with other techniques (Wierwille and Williges 1978), in order to better evaluate the operator-centred aspects of the workload system (see Table 1). Resources models (such as those discussed in Section 4) are useful for predicting task combinations which may lead to performance degradations due to resource competition effects but are not very helpful as models for subjective workload measures. At best they deal with the *attentiveness/attention* factor of Table 1. The principal underlying assumption of subjective measures is that humans possess a quantifiable internal representation of mental load (Revesman and Rokicki 1985). It might be expected that subjective workload measures will reflect experiential components such as those listed in the center column of Table 1, as well as subjective responses to the demands of the task (left column) and a perception of task performance (right column).

One main advantage that the subjective measure has over objective techniques is its high face validity as a measure of true (i.e. perceived) workload (Reid, Shingledecker and Eggemeier 1981, Skipper, Rieger and Wierwille 1986). The research on psychometric scaling by magnitude estimation (Pepermans and Corlett 1983, Stevens 1962) shows that humans are indeed able to quantitatively judge sensory magnitudes. To cite the frequently quoted assertion of Johannsen, Moray, Pew, Rasmussen, Sanders and Wickens (1979), 'If the person *feels* loaded and effortful, he *is* loaded and effortful whatever the behavioural and performance measures may show.' (p. 105). Moreover, many researchers (e.g., Wierwille and Williges 1978, Helm 1981) believe subjective measures to be the most promising workload assessment technique available at present. For these reasons, subjective measurement techniques are examined in detail in this report.

Subjective opinions have been solicited in the form of rating scales, questionnaires, and interviews. Questionnaires and interviews can provide valuable insights into mental load and related issues, however information can generally be extracted from rating scales in more specific detail than from questionnaires and interviews. The psychometric basis and the potential for more precise quantitative data renders rating scales a more promising candidate for mental load evaluation than questionnaires and interviews (Williges and Wierwille 1979). Therefore, this report focuses on rating scales as the principal media for subjective opinions.

There are several additional notable advantages of subjective measures. First, they have been found to be relatively sensitive to workload ranges below the threshold for overload (e.g., Casali and Wierwille 1984, Eggemeier 1984, Hicks and Wierwille 1979, Wierwille and Connor 1983). Further, they are considered to be nonintrusive by most researchers (e.g., Moray 1982, Reid, Shingledecker and Eggemeier 1981, Skipper et al. 1986), as the opinions are recorded before or after performance of the task or at selected breaks throughout, but not concurrently. Still the possibility of intrusiveness does exist, as the operator's awareness of his responsibility to accurately report the desired information may either introduce the additional task of self-monitoring of specific variables (feelings) or alter the manner in which the operator carries out the task.

Other benefits of employing subjective measures are the minimal instrumentation requirements (Skipper et al. 1986), the cost-efficiency (Casali and Wierwille 1983), the transferability to a wide range of tasks (Casali and Wierwille 1983) and to various stages of design, and the high degree of operator acceptance (Skipper et al. 1986). Several workload studies that were reviewed remarked explicitly on the ease with which operators could respond to the queries regarding subjective workload levels. There were no reported cases of studies being abandoned due to lack of operator understanding or acceptance, although this does not negate that possibility.

The principal drawback of subjective techniques is the uncertainty about what is actually being measured. For example, subjective measures may not distinguish effects due to cognitive task loading from those resulting from physical task loading (Moray 1982, Wierwille and Williges 1978). The problem of soliciting subjective estimates of mental workload is yet further confounded with differences in individuals' understanding of the concept of mental load. In his review of the subjective mental load literature, Moray (1982) criticises the activities of researchers by stating that '...little effort has been made explicitly to understand the origins of subjective feelings of load.' (p. 37). Helm (1981) supports this opinion. Since this critique, some investigators (e.g.,

Derrick 1981 and Derrick and Wickens 1984) have concerned themselves more directly with this aspect.

There are several other problems with subjective mental workload that are worth noting. Foremost among these is the poor diagnosticity of subjective measures (Eggemeier 1984). This deficiency stems, in part, from the inability to reliably map psychometric dimensions onto objective indicators of workload. The lack of diagnosticity of subjective measures suggests their utility as a screening implement, to determine the extent of loading. Attributing the mental loading to various aspects of the task would be left to the more diagnostic measures, such as secondary tasks.

A recent paper by Thornton (1985) on the *von Restorff* phenomenon in post-test workload ratings highlighted another potential problem with subjective opinions. Thornton concluded that the temporal location of a peak in workload affects the subjective perceptions in post-test ratings. The more proximal the workload peak was to the time of reporting, the more likely was the total task segment to be regarded as high workload. This memory-based effect has significant implications on how subjective mental load techniques should be administered and interpreted, and seems to support instantaneous measures as more reliable than the traditional *post hoc* measures.

In contrast to Thornton's finding, however, Yeh and Wickens (1985) did not find that variation in task loading within a task affected subjective ratings of workload. They concluded that perceived load is not a retrieval of information stored in working memory, rather '...it may reflect the experiences of an ongoing integration process.' (p. 765). It is possible that Yeh and Wickens did not vary the task difficulty sufficiently to cause the effects noted by Thornton.

Another factor that likely influences the ability of the operator to accurately reflect the level of mental loading is the degree of loading itself (Gartner and Murphy 1976). When the operator is concentrating intensely on the performance of a task, he is less aware of how hard he is cognitively working. This can be explained by the idea that his self-monitoring resources are being used to supplement overburdened resources dedicated to the task. In this situation, the operator will often report a low workload level, as he has been too involved in the performance of the task to be cognizant of the demands. This is consistent with the assumption that subjective workload is perceived (i.e. is at the level of consciousness) effort.

There are some (e.g., Williges and Wierwille 1979) who believe that rating scales developed for subjective assessments of mental load are usually situation-specific — they can not be generalised from the experimental conditions in which they were applied. In addition, the problem of rating biases exists in several forms (Wierwille and Williges 1978, Skipper et al. 1986), due to previous experience, changing impressions, emotional state, adaptivity, learning, central tendency, leniency, and natural abilities such as intelligence. Moray (1982) noted that a tendency exists for more intelligent subjects to rate tasks as harder than less intelligent subjects, which seems counter-intuitive. The explanation offered was that the more capable subjects may reflect on the difficulty of the task for a longer period than most.

Despite the drawbacks outlined above, subjective measures continue to provide relatively reliable estimates of mental load. The tremendous conceptual appeal of subjective measurement techniques underlies their merit for continued investigation and refinement.

6.2. Analytical Procedures

Procedures exist which utilise various mathematical formulations to model the workload situation. Some of the concepts that have been employed are optimal control theory (OCT— for an explanation of this, see Levison 1979), queuing theory (e.g., Senders 1983), classical control theory, and signal detection theory. However, perhaps the most widely used analytical methods are those ***task analytic*** procedures which are based on time-line analysis (Meister 1985, pp. 63-65), or some variation of this method.

Task analytic measures are concerned with the systematic evaluation of task demands. Techniques are available which attempt to examine task demands along dimensions of *effort* or *time* (Meister 1985). Arguably the most useful techniques operate in the time domain (Meister 1985, p. 77) although procedures based on an assessment of attentional demand have gained popularity in

recent years. Generally, time-based measures attempt to establish the *time available* indicator under the task demands heading of Table 1. This is generally accomplished by an information theoretic approach or by time summation procedures. The underlying assumption is that all task components require specific lengths of time to complete (Wierwille and Williges 1978). When the time available to the operator exceeds the time necessary to perform all components of the task, a spare capacity exists (within the *time* rather than the *effort* domain). However, if the time required by the task is greater than the time available, the operator is assumed to be overloaded.

A variety of methods has been developed and implemented to assess the time requirements of various task combinations. Wierwille and Williges (1978) provide a list of some of the more prominent techniques. Time-line analysis, which breaks a job down into its elemental components of behaviour (to a level which is appropriate to the analysis or is limited by our knowledge of the job requirements) and associates a time value to each element, is the most widely used analytic technique. The standard technique possesses a fundamental flaw, it is generally assumed that all elements of the task are performed serially. Stated in other terms, '...it normally assumes a simplified 'single-channel' model of human attention in which any observable activity demands full attention.' (Wickens 1984). In fact, many investigations have evidenced this to be untrue. A task analytic method that attempts to overcome this problem deserves some mention here. The Function Interlace Technique (FIT), developed by Wingert (1973), attempts to account for the operator's timesharing abilities.

Wingert developed the function interlace technique to account for the ability of skilled operators to apparently perform some functions involving multiple tasks more efficiently than a single channel hypothesis of human information processing would allow. In Wingert's words,

> 'The term "function interlace" was coined to describe the phenomenon most easily understood as the apparent ability of skilled operators to perform tasks simultaneously... Function interlace may in reality stem from our inability to model man's behaviour at as molecular a level as we must to account for his capability.' (p. 376).

The FIT assumes that there is a certain proportion of a task that can be performed while the operator is simultaneously engaged in another task. The extent to which this can be accomplished depends on the specific task components, or the combination of input mechanisms and output mechanisms involved. The input mechanisms considered are: (1) visual; (2) auditory; and (3) kinaesthetic — responses are, (1) motor; (2) vocal; and (3) none. Each input in combination with each output makes up nine task components in total. Wingert attached interlace coefficients to component pairs based on simulation studies and various analytical techniques. For example, Wingert determined that a task component that involves visual input and vocal output can be timeshared with a task component involving auditory input and motor output with eighty percent efficiency (i.e. an interlace estimate of 80%). By analysing different functions (such as systems monitoring, communication, flight control etc.) with respect to the components involved, Wingert arrived at function interlace coefficients that predict the extent to which various high level activities can be timeshared effectively.

The concept of function interlacing ties in closely with Wickens' view of multiple resources. Wingert's input and output mechanisms are analogous to Wickens' input and output modalities, although Wingert expands his technique to include tactile, olfactory, and gustatory input sensations as well as several motor channels of output and an output mechanism devoted to controlling the direction of visual gaze. Each of these channels can be viewed as potentially distinct modalities through which information is transmitted by the human operator. The rationale for the Wickens' multiple resources model comes from the observed interference resulting when certain tasks are performed concurrently (Derrick and Wickens 1984). This concept of interference is embedded in the interlacing technique.

Wingert attached workload estimates to individual tasks by a straightforward time component summation method, then applied the interlace technique as follows. If the workload of Task 1 is W_1 and that of Task 2 is W_2, and the function interlace coefficient of Tasks 1 and 2 is

I, then the equation to find the overall workload ($\sum W$) of tasks 1 and 2 when performed concurrently is:

$$\sum W = W_1 + W_2 (1-I).$$

In this manner the overall workload for a man-machine system which involves the performance of concurrent tasks can be estimated, without the conservatism inherent in the traditional time-line analysis.

The limitations of the technique are due to the inability to attach accurate interlacing coefficients to the various task component pairs or to the tasks themselves. Until these values can be determined with some reliability, the functional interlace technique remains conceptually appealing but limited in application.

6.3. Physiological Correlates

A vast body of literature deals with the measurement of mental load by one or more physiological parameters. Physiological measures are distinguished from other measures by their focus on involuntary changes in the operator state. The underlying principle of physiological measures is stated in the following passage, taken from Wierwille (1979):

> 'As operator workload changes, involuntary changes take place in the physiological processes of the human body (body chemistry, nervous system activity, circulatory or respiratory activity, etc.) Consequently, workload may be assessed by the measurement and processing of the appropriate physiological variables.' (p. 575).

Under this assumption, the task becomes one of determining which physiological variables are appropriate to measure the workload in any specific situation.

The definition of mental load from a physiological point of view (see Section 3) refers to the ***activation of the organism,*** which is the operator's level of arousal or excitedness. In order to provide a reliable measure of mental load, the parameters measuring the activation level must be able to isolate effects resulting from mental load from those due to a host of other possible contributors, including for example, physical load and cognition not related to the imposed task. As yet, this has not been accomplished in a reliable fashion.

Wierwille (1979) published a review of the 14 most common physiological measurement techniques (listed in Figure 5). The conclusion reached (see also Singleton 1971) is that no single physiological technique is able to provide valid measures of mental load, however additional information is gained by employing physiological measures in conjunction with behavioural measures. The most promising techniques noted are pupil dilation, evoked cortical potentials, and body fluid analysis.

7. SUBJECTIVE WORKLOAD METRICS

In this Section, some of the most used subjective workload metrics are discussed. In the last 5 years, one workload metric has dominated the area of subjective workload assessment. This metric, the Subjective Workload Assessment Technique (SWAT), has seen use in a wide variety of situations (see Sections 7.1.4, 7.1.6 and 7.1.7) and has undergone extensive reliability and validation testing. A contender is the NASA Task Load Index (Section 7.2), although the older Cooper-Harper (CH) and Modified Cooper-Harper (MCH) scales have appeal because of their ease of

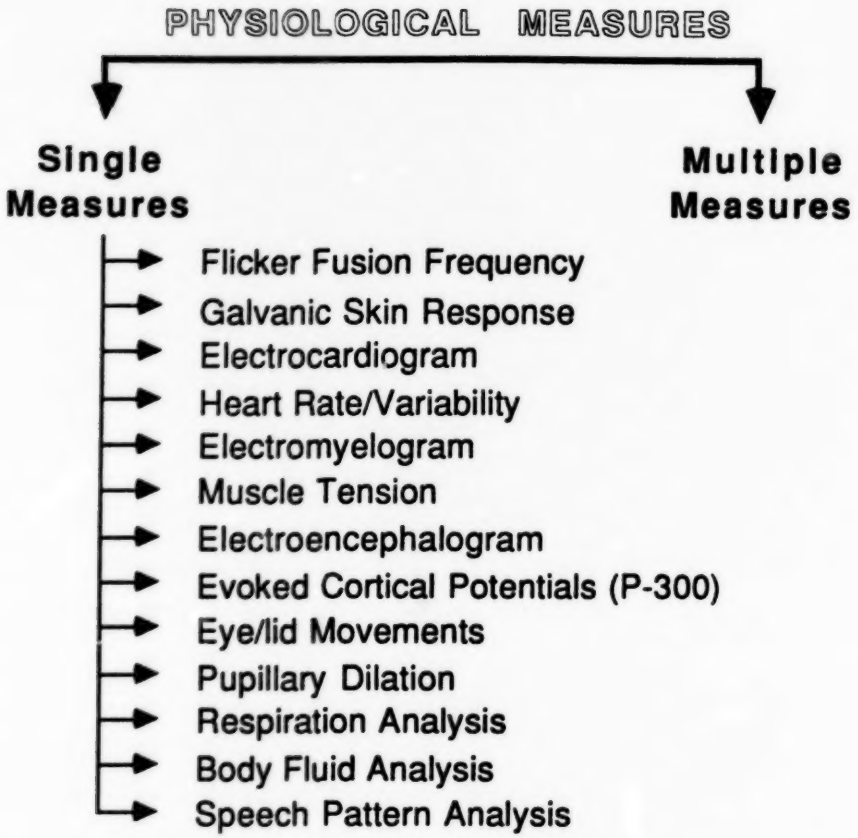

Figure 5. *Physiological Measures of Mental Load*

administration (Section 7.3). Finally in this Section, the application of the psychometric technique of ***magnitude estimation*** to subjective workload studies is discussed (Section 7.4).

7.1. The Subjective Workload Assessment Technique (SWAT)

SWAT, developed by Reid, Shingledecker and Eggemeier (1981), is likely the most widely used subjective technique today. SWAT was developed '...to provide a workload rating scale with known metric properties and with the potential for generalised applicability.' (Eggemeier 1984, p. C-13). Despite its popularity, SWAT does not escape criticisms concerning its validity. The technique is now undergoing its sixth year of validation, and there are many aspects of SWAT that are yet to be adequately addressed. An examination of SWAT is presented below, accompanied by a discussion of its advantages, limitations, validity and applicability. In addition, steps for the further validation of SWAT are outlined.

SWAT recognises the multidimensionality of workload. Specifically, the technique forms a workload index by combining subjective estimates of *mental effort load*, *time load*, and *psychological stress load*. Within each dimension, ratings can take the values 1,2 or 3. The administration of SWAT takes place in two phases: scale development and event scoring.

TABLE 2: SWAT SCALE DESCRIPTORS

Level	Descriptor
	Mental Effort Load
1	Very little conscious mental effort or concentration required. Activity is almost automatic, requiring little or no attention.
2	Moderate conscious mental effort or concentration required. Complexity of activity is moderately high due to uncertainty, unpredicability, or unfamiliarity. Considerable attention required.
3	Extensive mental effort and concentration are necessary. Very complex activity requiring total attention.
	Time Load
1	Often have spare time. Interruptions or overlap among activities occur infrequently or not at all.
2	Occasionally have spare time. Interruptions or overlap among activities occur frequently.
3	Almost never have spare time. Interruptions or overlap among activities are very frequent, or occur all the time.
	Psychological Stress Load
1	Little confusion, risk, frustration, or anxiety exists and can be easily accommodated.
2	Moderate stress due to confusion, frustration, or anxiety noticeably adds to workload. Significant compensation is required to maintain adequate performance.
3	High to very intense stress due to confusion, frustration, or anxiety. High to extreme determination and self control required.

7.1.1. Dimensions of SWAT

Mental Effort Load

Mental effort, as described to subjects, deals with concepts such as planning, task complexity, automaticity, and attention. The extent of the relationships between each of these and mental effort, or among these and mental load, is not well known. Further, the subjects' interpretations of these concepts is crucial to their subjective response. SWAT descriptors for each level of mental effort load are shown in Table 2.

Mental effort is apparently closely related to mental workload. In fact, many authors use the two terms interchangeably.

Time Load

Time load refers to the fraction of the total time occupied in performing the task. This dimension is described in terms of interruptions, spare time and the requirement to timeshare tasks. The applicability of the time dimension to different workload situations is unclear, as multiple descriptions of time-related concepts are associated with each level (see Table 2). For example, the highest level on the time load scale refers both to a lack of spare time as well as to the frequent occurrence of simultaneous tasks. This description would not be suitable for a single task that imposes heavy time pressure, as the simultaneous task feature is not present, although the task might otherwise be deserving of the highest time load rating.

Time load also appears to be closely related to workload, in that both parameters seem to be associated strongly with task performance although the relationships are not simple. Wickens cites studies which indicate that subjective estimates of workload are relatively more sensitive to the number of competing activities (time load) while primary task performance is apparently more sensitive to the difficulty of a single task (Wickens 1984, p. 328). Time is often the crucial parameter in predicting the performance of complex man-machine systems, and likely should not be excluded from any performance evaluation.

Psychological Stress Load

Terms referred to in describing this dimension are confusion, frustration, and anxiety as shown in Table 2. These items are all listed in Table 1 as potential workload indicators in the operator-centred and performance components of the workload system. Stress likely modulates the level of mental load determined by time and effort loads. This would indicate that psychological stress is not an appropriate dimension of workload in a simple linear combination with time and effort loads.

7.1.2. SWAT Scale development

The purpose of the scale development stage in SWAT, is to create a combination rule that maps *effort*, *time load*, and *stress* onto *workload*. This phase requires the subjects to rank, in order of increasing workload, all 27 possible combinations (3 dimensions, 3 levels each) of the time, effort, and stress levels. Verbal descriptions of the various combinations are normally presented on cards. It is usual in this process to instruct the subjects to imagine a situation that can be described by each of the 27 cards. The subjects are advised to first sort, into three groups, the cards corresponding to low, moderate, and high workloads. Then pairs of cards are ordered within groups by comparing the workloads of the imagined events that correspond to the description on the cards. The subjects are instructed to divorce themselves from any one particular task environment when ordering the cards. However, the extent to which this can be successfully carried out depends on the experience of the subject. A group of pilots, for example, will probably compare various flight experiences when performing the card sort.

The degree of agreement of the rankings among the subjects can be determined using Kendall's coefficient. If the result is greater than some criterion value, good agreement would be

assumed and the subjects would be treated as a single group. Below this level, however, it would be assumed that subjects are using different ranking strategies. Therefore, subjects should be classified into one of three groups according to whether time, effort, or stress appears to be the dominant dimension in determining the ranks. Unique workload composition rules can be formed for each group (a detailed explanation of this procedure is provided by Reid, Eggemeier and Nygren, 1982). This method of dealing with individual differences introduces some concern about SWAT. As individuals evidently carry various views as to what factors are dominant in determining workload, the validity of comparing or averaging the resulting undimensional workload estimates across groups is questionable.

A psychometric scaling technique known as conjoint analysis (Nygren 1982) folds the three dimensions of *effort*, *time*, and *stress*, onto a single dimension which is assumed to represent overall mental workload. This is achieved by performing a series of axiom tests to identify the combination rule to be applied (these tests are described in Reid, Shingledecker and Eggemeier 1981) and then applying the scaling procedure. An advantage claimed for the use of conjoint analysis in the analysis of SWAT ratings, is that ordinal data from subjective opinions are translated onto a 27 point workload scale with interval scale properties.

Reid, Shingledecker, and Eggemeier (1981) refer to reliability checks on the card sorts, administered as follows. Nine card combinations, are presented to the subjects for comparison. Three of the pairs have very few steps between them in the workload index, three have a moderate number, and three represent workload ratings at opposite ends of the scale. They find that 'Reliability checks within approximately two weeks of event scoring have indicated that pair comparison responses agree with original rankings approximately 87% of the time...' (p. 524). Further, a sample size of eight pilots showed 75% agreement after three months. At first glance, these results appear admirable. However, it is highly probable that most disagreement between the original rankings and the paired comparisons would occur for small differences, rather than for moderate differences or for the pairs on opposite ends of the workload index. If the test-retest differences were in the small intervals alone, then the reliability for these differences may be quite low. Reliability of small differences is important if the scale is to have high discriminablity.

7.1.3. Event scoring

In the latter phase of SWAT administration, subjects score actual events or workload situations on each of the three ordinal scales corresponding to the dimensions of the metric. A SWAT rating of a specific workload situation is converted to one of the 27 points on the workload scale which quantifies the overall mental load. The results obtained from the event scoring are used directly in the comparisons of various workload situations.

7.1.4. Applicability of SWAT

Various tasks have been employed to assess the range of application of the SWAT technique, with generally positive results. SWAT has proven itself to be sensitive to workload differences (as indicated by objective parameters of task demands) present in the following types of tasks:

(1) tracking (Reid, Shingledecker and Eggemeier 1981, Vidulich and Tsang 1985);

(2) short-term memory (Eggemeier, Crabtree, Zingg, Reid and Shingledecker 1982, Eggemeier, Crabtree and LaPointe 1983);

(3) spatial transformation (Vidulich and Tsang 1985);

(4) spatial memory (Eggemeier and Stadler 1984);

(5) display monitoring (Notestine 1984); and

(6) multi-faceted tasks of perception, central processing, and motor response (Crabtree, Bateman and Acton 1984).

It is worthwhile to note that performance measures for several of the tasks were more sensitive to changes in workload than were the SWAT ratings. Tasks which showed this effect were the Interval Production Task (IPT) (Crabtree, Bateman and Acton 1984), tracking (Vidulich and Tsang 1985), and response accuracy on the monitoring task (Notestine 1984). In view of previous argument (see Section 6.1.1), this doesn't necessarily state a case for the superiority of objective measures over subjective measures but rather underlines the importance of including appropriate objective measures in a workload test battery, if this is possible.

Although tasks such as those above are all performed by operators of aircraft systems, they only partially represent the full range of tasks that may be performed in an aviation environment. A comprehensive set of operator behaviours, consisting of processes and activities, is provided by Wierwille and Williges (1978). Full validation of the applicability of SWAT as a global workload measure, that can attach workload values to activities in an aviation environment, will involve sensitivity testing for all appropriate behaviours.

7.1.5. Validity of SWAT

The component dimensions of SWAT (see Boyd 1983) appear to come more from an intuitive approach than from a rigorous empirical derivation. In view of this background a principal assumption usually invoked in SWAT scale development, i.e. that all three estimated dimensions are subjectively orthogonal (Reid, Shingledecker and Eggemeier 1981), should be tested. Boyd (1983) attempted to test the validity of this assumption through the selection of specific tasks, and the manipulation of certain parameters, which were intended to affect the subjective dimensions of time, effort, and stress independently. It does not appear that Boyd was able to successfully control each dimension independently. For example, to manipulate the time load, several interruptive messages were sent to the subject's video screen. These messages likely affected the stress dimension and possibly effort, as well as time load. Similar problems accompanied the attempts to separately manipulate effort and stress. As this explicit attempt to independently manipulate time, stress, and effort failed, the contention that there are indeed correlations among the subjective dimensions cannot be dismissed.

SWAT was originally designed for subjects to submit ratings immediately following completion of the task, or following the completion of individual phases of the task. Due to its apparent universality and applicability across a range of tasks, SWAT's uses have expanded to include both projective (Eggleston and Quinn 1984, Reid, Shingledecker, Hockenberger and Quinn 1984) and reflective measures (Arbak, Shew and Simons 1984).

7.1.6. Projective SWAT

One advantage that a subjective technique such as SWAT might provide is the potential to be used to solicit workload estimates in the preliminary stages of system design. The use of predictive workload metrics in this way, may avoid the burden of manufacturing equipment of alternative designs in order to compare their respective workloads in full scale tests or detailed simulations. Reid et al. (1984) have termed this procedure *Projective SWAT* (or Pro-SWAT). A study which has examined the feasibility and reliability of Pro-SWAT (Eggleston and Quinn 1984) provides the following description:

> '...the Projective Subjective Workload Assessment Technique,... utilises written and verbal descriptions of both a system concept...and a task (mission) to be completed...Pro-SWAT requires task-knowledgeable raters to mentally project themselves into the operation of the defined system, imagine performing the task, and then report the magnitude of the workload "experienced" at selected times...' (p. 695).

It is evident that great care must be taken in the design of a Pro-SWAT study, in order to obtain measures that may reliably reflect the workload of the tasks described. Its primary limitations are in the ability of the subjects to accurately assess workload based solely on task and equipment descriptions and, of course, on the adequacy of the descriptions themselves. The workload ratings likely vary with the subjects' knowledge of the system being rated and therefore there is a tradeoff between investigator bias and rater bias, depending on whether detailed or vague descriptions are used.

Eggleston and Quinn (1984) suggest several practical methods for reducing rating variance and increasing the validity of workload estimates for Pro-SWAT. Further investigation is necessary to determine the value of Pro-SWAT workload comparisons.

7.1.7. Reflective SWAT

Reflective SWAT was designed as a tool for gathering workload ratings through the operator's reflections, often after a considerable time delay, on mission performance. Arbak et al. (1984) see the following advantages to this approach: (1) *post hoc* estimates of workload reflect more general conditions; and (2) estimates can be obtained where immediate workload ratings are unavailable. It is recommended that a reflective SWAT be employed along with an interview.

Although subjects do actually experience the workload before the reflective report, many of the limitations noted in Pro-SWAT apply here, particularly those due to event descriptions. In reflective SWAT, it is generally the responsibility of each subject to define mission segments (e.g., according to weather conditions, task, or terrain). Thus often the segments rated are not the same among subjects.

Reflective SWAT relies on the ability of the subjects to remember the workload experienced over delayed reporting intervals. Two studies that have examined this phenomenon (Eggemeier, Crabtree and LaPointe 1983, Eggemeier, Melville and Crabtree 1984) showed no significant difference between means of immediate and delayed workload estimates using SWAT, although differences were noted within subjects. In addition, Eggemeier, Melville and Crabtree determined that an intervening task between the primary task and the subjective report caused no significant difference in workload ratings. This second conclusion requires verification under conditions where the relationship between the primary and intervening tasks is varied.

The benefits of reflective SWAT claimed by Arbak et al. (1984) have not been validated. Again, further investigation must determine the applicability of Reflective SWAT to the evaluation of workload.

7.1.8. Further validation of SWAT

In order to validate SWAT as a mental workload metric, empirical evidence must be gathered that points to *effort*, *time*, and *stress* as the sole (or at least the principal) dimensions that constitute mental load. If correlations among the SWAT dimensions are observed in empirical studies, then the effects of these on the combination rule(s) and workload ratings must be determined. Theoretically, the additive conjoint technique is not applicable when the variables are correlated. Thus, another combination rule, using perhaps the multiplicative or distributive versions of the conjoint analysis may be appropriate. Or perhaps the effect of the correlations is not so significant as to cause notable perturbations of the workload estimates obtained by the additive procedure. In

either case, sensitivity analyses on the scale development phase and event scores should be performed over the ranges of correlations observed.

Due to the poor reproduceability of the card sort within subjects, the sensitivity to small and large variations in rank ordering should be assessed. This examination would encompass the following steps:

Step 1: Determine an initial ranking of the 27 matrix cells.

Step 2: Analyse the properties of the configured data in order to determine the appropriate conjoint technique to apply.

Step 3: Perform the conjoint analysis to develop the scaling factors of the combination model.

Step 4: Provide event scores — either means of actual data or fictitious values — and derive the overall workload score.

Step 5: Reconfigure the card sort, allowing between 0% and X% discrepancies among small or large intervals from the original data (X% is the highest observed error rate in small or large intervals of the reliability checks), and recalculate the scaling factors.

Step 6: Perform a Monte Carlo simulation of Steps 2 through 5, drawing from a distribution determined from historical data.

The above measures for validating SWAT can form part of an investigation to assess the applicability of SWAT across operator behaviours, tasks, and various workload situations. With the additional knowledge that can be gained from this evaluation, SWAT can be either discarded as a workload metric or, more likely, employed strategically with an understanding of the limitations of the procedure.

7.2. NASA Task Load Index (NASA-TLX)

The NASA Task Load Index (an earlier version of the scale was called the *NASA Bipolar Rating Scale*) is a recently developed subjective workload assessment technique (Hart, Battiste and Lester 1984), which has the operator estimate 6 aspects (9 in the case of the NASA Bipolar version) which are believed to be related to workload, *viz.*: mental demands, time demands, own performance, frustration level, effort and physical demands. In a scale development phase, subjects determine the extent to which each item contributes to their overall perception of workload in a series of paired comparison judgements. Weightings for each scale are calculated, usually for each subject involved in the assessment. Events are scored by recording an estimate, on each of the six scales, as a mark on a 120 mm line. The line is divided into 20 equal intervals by vertical tick marks. Verbal bipolar descriptors anchor each end of the scales. The workload value of the event is then calculated by multiplying the individual scale values by the corresponding weighting coefficient. Scale descriptors are shown in Table 3 and an example of a rating sheet for the NASA TLX is shown in Figure 6.

Vidulich and Tsang (1985) compared the NASA Bipolar technique to SWAT in a dual-task paradigm (tracking and transformation). Both techniques were equally sensitive to the various workload situations. As well, Vidulich and Tsang concluded the following: 'There does not seem to be any large distinction between the SWAT and NASA weighted-bipolar validity; both appear to be fairly valid. Therefore, the use of either technique as an operational tool seems justified.' (p. 75).

TABLE 3: NASA-TLX DESCRIPTORS

RATING SCALE DEFINITIONS		
Scale Title	Endpoints	Descriptor
MENTAL DEMAND	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

7.3. Cooper-Harper Scale

The Cooper-Harper (CH) rating scale (Cooper and Harper 1969) has been widely accepted in the aviation industry as a reliable indicator of aircraft handling qualities. However, researchers have encountered problems when attempting to apply the CH scale to other than aircraft handling problems. A modified version of the CH, the Modified Cooper-Harper scale (MCH), has been developed (Wierwille and Casali 1983). The intention of this development was to create a globally sensitive indicator of mental load. The MCH, which incorporates a decision-tree structure (see Figure 7), has been shown to be sensitive in a number of workload situations (Skipper et al. 1986) although much controversy exists over its applicability outside the originally intended domain of flight. The MCH scale has a very high level of operator acceptance and it is extremely easy to implement.

7.4. Magnitude Estimation

Magnitude estimation is a widely used psychometric technique for eliciting subjective information on the magnitudes of various sensory stimuli. It has been found (Pepermans and Corlett 1983, Stevens 1962) that for prothetic stimuli (i.e. for stimuli that may vary in magnitude along a dimension) a power law exists of the following form:

$$R = k\,(S - S_0)^n,$$

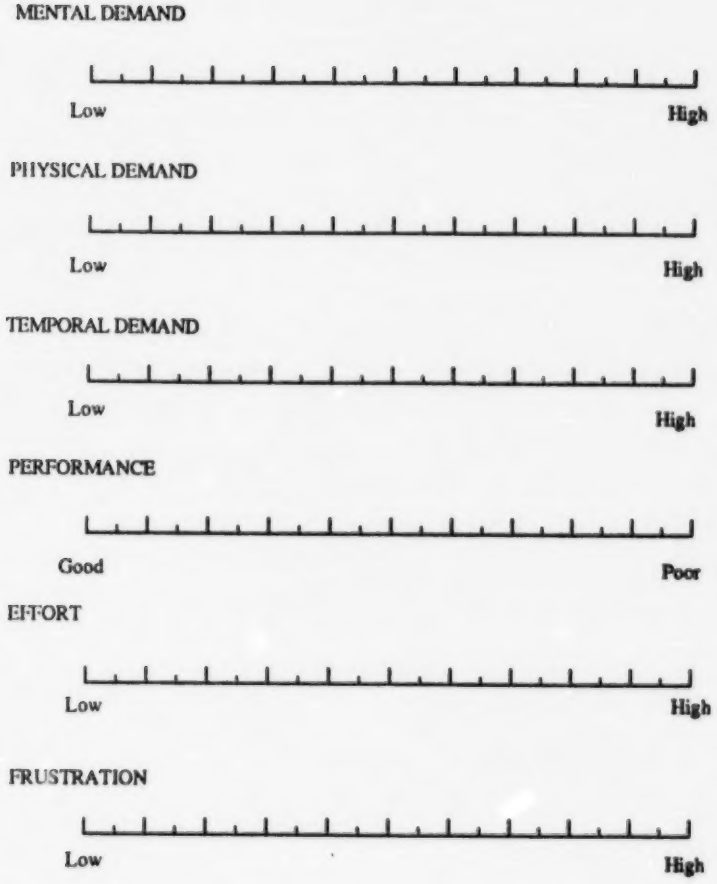

Figure 6. *NASA Task Load Index (NASA-TLX) Rating Scales*

where, R is the magnitude of the sensation in the psychological domain;
S is stimulus magnitude;
S_0 is the threshold of perception for the stimulus;
n is a modality-specific exponent; and
k is a constant.

This relation can be applied to subjective mental load, under the assumption that workload can be treated as if arising from prothetic stimuli. If a subject is asked to estimate the mental load of a specific workload situation and the response can be plotted against a workload indicator (perhaps a combination of objective task parameters) on a power curve, then we might assume

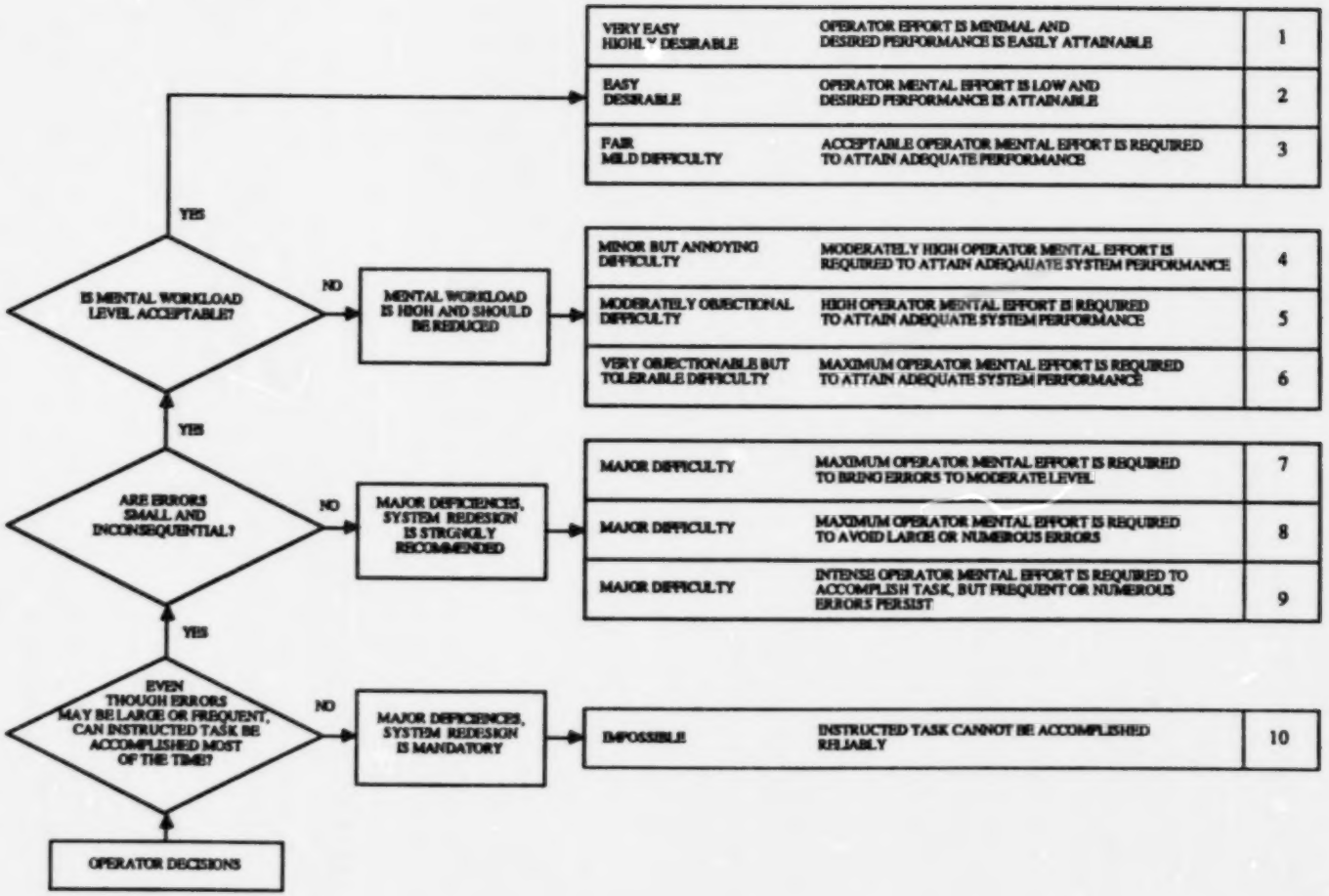
OPERATOR DECISIONS
EVEN THOUGH ERRORS MAY BE LARGE OR FREQUENT, CAN INSTRUCTED TASK BE ACCOMPLISHED MOST OF THE TIME?
NO
MAJOR DEFICIENCIES, SYSTEM REDESIGN IS MANDATORY
IMPOSSIBLE
INSTRUCTED TASK CANNOT BE ACCOMPLISHED RELIABLY
10
YES
ARE ERRORS SMALL AND INCONSEQUENTIAL?
NO
MAJOR DEFICIENCIES, SYSTEM REDESIGN IS STRONGLY RECOMMENDED
MAJOR DIFFICULTY
MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO BRING ERRORS TO MODERATE LEVEL
7
MAJOR DIFFICULTY
MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO AVOID LARGE OR NUMEROUS ERRORS
8
MAJOR DIFFICULTY
INTENSE OPERATOR MENTAL EFFORT IS REQUIRED TO ACCOMPLISH TASK, BUT FREQUENT OR NUMEROUS ERRORS PERSIST
9
YES
IS MENTAL WORKLOAD LEVEL ACCEPTABLE?
NO
MENTAL WORKLOAD IS HIGH AND SHOULD BE REDUCED
MINOR BUT ANNOYING DIFFICULTY
MODERATELY HIGH OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE
4
MODERATELY OBJECTIONAL DIFFICULTY
HIGH OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE
5
VERY OBJECTIONABLE BUT TOLERABLE DIFFICULTY
MAXIMUM OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE SYSTEM PERFORMANCE
6
YES
FAIR MILD DIFFICULTY
ACCEPTABLE OPERATOR MENTAL EFFORT IS REQUIRED TO ATTAIN ADEQUATE PERFORMANCE
3
EASY DESIRABLE
OPERATOR MENTAL EFFORT IS LOW AND DESIRED PERFORMANCE IS ATTAINABLE
2
VERY EASY HIGHLY DESIRABLE
OPERATOR EFFORT IS MINIMAL AND DESIRED PERFORMANCE IS EASILY ATTAINABLE
1

Figure 7. *The modified Cooper-Harper (MCH) Workload Scale*

with reasonable certainty that the subject is estimating the magnitude of that workload indicator.

Gopher and Braune (1984) attempted to predict the amount of cognitive resources invested in 21 experimental conditions using the magnitude estimation technique. The subjects performed a reference task, to which the value 100 was arbitrarily assigned. The estimates from all other conditions were normalised for each subject, by the largest numerical value used, in order to calculate inter-subject statistics. The tasks varied in terms of the resources required *viz.*: perceptual-motor control, short-term memory, verbal and spatial abilities, selected and divided attention, and timesharing capabilities. They found that:

> '...consistent profiles of the perceived load of tasks can be obtained despite the freedom given to subjects in the selection of values and logic of rating, the multitude of tasks, and their vast heterogeneity. In addition, a consistent differentiation was found between single- and dual-task conditions, and there was a clear indication of practice effects ' (p. 525).

Normally in psychometric scaling experiments, the modality-specific exponent (n) and the constant (k) can be derived from the subjective estimates and the magnitude of the known physical dimension that drives sensory experience. However, experiments attempting to measure workload must infer the variable(s) that drives experience and causes the subject to perceive mental load. Gopher and Braune (1984) assume that a common function (they hypothesize the function ***invested resources***) is driving the subjective estimates across all tasks and thus they derive a single value for n and k for all 21 experimental conditions. The values of n and k were calculated by arbitarily assigning resource values of 1 and 2 units to Sternberg auditory-verbal memory search tasks with set sizes of 2 and 4 respectively (information contents of 1 and 2 bits). From these assumed values, the simultaneous solution to two equations yields the values of the two unknowns.

The validity of this procedure rests on the belief that a single function, say processing resources, was estimated by all subjects for all tasks. It is conceivable that various physical quantities are driving the subjects' responses and that n and k are task-specific. Thus, more data must be collected in order to establish the constancy, or lack of it, in the derived values for n and k over tasks.

Magnitude estimation may help to unfold commonalities among various tasks and to find relationships among different workload parameters. Using the concept that underlies Cross-Modality Matching (CMM) research, one can '...compare changes in two variables from different sense modalities with each other, if the perceived sensation they cause is measured in the same third sense modality' (Pepermans and Corlett 1983, p. 171). This can apply to perceived load as follows: if the manipulation of workload parameter A causes the same change in perceived load as that of workload parameter B, these manipulations can be assumed to cause identical workload demands on the subject. In this manner, correlations among various workload indices that have been proposed in the literature could be determined and the relationship between mental workload and human performance may become apparent. Thus, magnitude estimation may prove to be auspicious, in light of its sound scientific background.

8. DISCUSSION

It is evident that a large degree of controversy still exists over mental workload. What is it? How can it be modelled? How can it be measured? What is the best approach to take when investigating workload? This report attempts to develop an appropriate framework from which experimentation might provide answers to some of these questions.

The principal advance that has come out of the mental workload research over the past decade is the conclusion that mental load is a multidimensional phenomenon. From this emerges the assertion that no one metric will suffice as a global measure of mental load. A battery of techniques should be employed so as to gain a maximum amount of knowledge about the workload situation. More diagnostic measures should be employed in tandem with more sensitive measures. Techniques that are sensitive to overloads should be used with those that can discriminate

workload levels below threshold. Objective measures should be combined with subjective estimates. Workload parameters which represent each component — task inputs, operator-centred variables, and response outputs — should be present in an evaluation of mental load. Finally, methods that are sensitive to various operator behaviours should be employed together in multi-task environments.

The utility of describing mental load as a scalar quantity is therefore open to question. Perhaps a vector representation which maintains the dimensions used to derive the workload estimate should be employed. Thus, comparisons of alternative designs and workload situations require the interpretation of workload vectors, as opposed to simply choosing the one with the larger overall workload value. A scalar approach assumes that various tasks, differing in the nature and source of their ***workload***, can be mapped onto the single derived workload dimension (Helm 1981). A comparison of systems on the basis of individual scores, for each of the dimensions rated, would be a simple implementation of a vector approach.

There is a vast number of potential workload indicators that have been and are being investigated. The measurement of each parameter may contribute to the knowledge base of mental load, but as Gopher and Braune (1984) point out,

> 'It should be recognised that originally, the main theoretical justification for instituting the workload concept was to attempt to improve the ability to predict performance, given task conditions. What is the sense and what is the value of developing a workload measure that does not correspond to (or is only weakly related to) the actual behaviour of subjects?' (p. 521).

Although methods and techniques designed to predict mental load or human performance in complex man-machine systems are far from perfect, the knowledge that has been gained from research in this area thus far, is advantageous to the human engineer who needs to evaluate system performance. The work done by Aldrich, Craddock and McCracken (1984) is a good example of a ***hands on*** approach to workload measurement. The question of interest for these authors was whether to configure a proposed light battlefield helicopter design for one or two crewmembers. Twenty-nine separate scout and attack mission segments were analysed on a timeline to determine exactly what the crewmembers were doing throughout the entire mission. The functions were classified into critical flight control, support, and mission. For each performance element, estimates of sensory, cognitive, and psychomotor workloads/durations were obtained. The goal was to determine points of excessive workload and the workload density for each segment.

Using computer algorithms, Aldrich et al. (1984) modelled the mission from both a one and two crewmember perspective, using an established workload threshold as a benchmark. Despite the numerous assumptions on which the study is based, and in spite of the approximations and estimations required in order to attach workload values to mission segments, the study employs a systematic and conceptually appealing method of comparing various system design proposals. This provides an excellent alternative to purely qualitative assessments during the early stages of design, and an alternative to full manned system simulation.

Most of the workload measurement techniques which have been discussed thus far, are aimed at determining the cumulative mental load of an operator performing a task. In fact workload is variable over time, just as the task demands and the operator's capacity may also vary with time. Often the investigator is interested in detecting any or all specific points of overload. Here instantaneous measures of workload (IMWL) are of value.

One study which investigated the feasibility of IMWL was conducted by Antin and Wierwille (1984). The following types of measures were employed: primary task performance; secondary task performance; physiological (pulse and respiration rates); and subjective opinion. Results indicated that the primary task performance and the subjective opinions showed great promise, although much research remains to be done.

Subjective opinions remain the most promising workload measurement technique at the moment, a strong virtue being the relative sensitivity below overload threshold values. Yet there

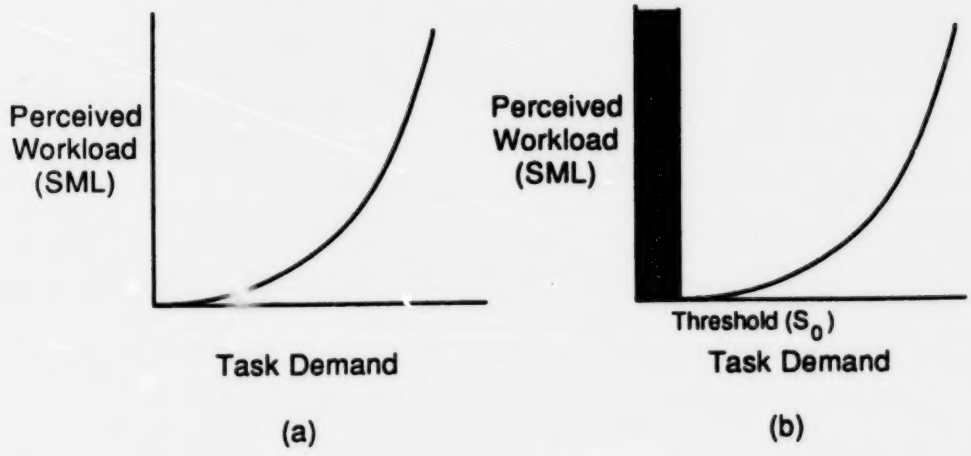

Figure 8. *Perceived Load-Workload Relationships*

is no technique available which is sensitive to workload ranges well below threshold. This can be explained in two ways:

(1) Perceived mental load maps onto actual workload as depicted in Figure 8a. The zero objective load point is perceived as zero workload ; as objective load increases, perceived workload increases with it (*cf.* Stevens' Law $R = kS^a$).

(2) Perceived mental load maps onto actual workload as depicted in Figure 8b. Workload is not perceived until a certain threshold value of objective load is exceeded (not to be confused with the overload threshold). Above this value, perceived workload behaves as in (1). This explanation ties in with the *response model* of strain described by Wiener (1982). Mental load is perceived when it is significant enough to cause a cognitive strain. Above this level, the perceived load reflects the mental strain of the subject (*cf.* Stevens' Law $R = k(S - S_0)^a$).

A sound theoretical explanation of subjective workload may carry with it the solution to this uncertainty. At present, this problem remains just one of the many points of controversy surrounding the measurement of mental workload.

9. CONCLUSIONS

The motivation for this work in the area of workload assessment and measurement is found in the need for properly validated, and in the long term standardized, human engineering methods which can be used by the Canadian Forces and Canadian Industry. Workload assessment seems to be a logical starting point for a human engineering tool kit, in light of the dramatic increase in the information presentation capabilities of contempory military equipment and the resultant possibility of perceptual and cognitive overload. It is concluded that further research on both a theoretical

and empirical level is needed in order to develop a comprehensive knowledge base on which informed decisions can be based. Specific areas for directed research include:

(1) The applicability of promising techniques to various workload situations should be determined through empirical study. These must account for individual differences in operator capacities, strategies, and selected performance criteria, along with the many operator behaviours that exist in a flight environment.

(2) The psychometric technique of magnitude estimation should be validated, and subsequently used to determine relationships among subjective parameters and indices of workload such as SWAT and NASA Bipolar. These can simultaneously be related to objective workload parameters in an effort to bring together the various divergent approaches to mental workload measurement that now exist.

(3) The limitations of various measurement techniques should be established through reliability and sensitivity analyses.

ACKNOWLEDGEMENTS

The authors are grateful for the information provided by Dr. Neville Moray of the University of Toronto and Mr. Gary Reid of the Armstrong Aerospace Medical Research Laboratories.

References

AGARD-CPP-216, 1977, *Methods to Assess Workload* (Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development).

AGARD-CPP-217, 1977, *Studies on Pilot Workload* (Neuilly-sur-Seine, France: Advisory Group for Aerospace Research and Development).

ALDRICH, T. B., CRADDOCK, W., and MCCRACKEN, J. H., 1984, A computer analysis to predict crew workload during LHX scout-attack missions Volume 1. MDA903-81-C-0504/ASI479-054-I-84(B), U.S. Army Research Institute Field Unit, Fort Rucker, Alabama.

ALLPORT, D. A., ANTONIS, B., and REYNOLDS, P., 1972, On the division of attention: a disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology*, **24**, 225-235.

ANTIN, J. F., and WIERWILLE, W. W., 1984, Instantaneous measures of mental workload: An initial investigation. In *Proceedings of the Human Factors Society's 28th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 6-10.

ARBAK, C. J., SHEW, R. L., and SIMONS, J. C., 1984, The use of reflective SWAT for workload assessment. In *Proceedings of the Human Factors Society's 28th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 959-962.

BOYD, S. P., 1983, Assessing the validity of SWAT as a workload measurement instrument. In *Proceedings of the Human Factors Society's 27th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 124-128.

CASALI, J. G., and WIERWILLE, W. W., 1983, A comparison of rating scale, secondary-task, physiological, and primary task workload estimation techniques in a simulated flight task emphasising communications load. *Human Factors*, **25**, 623-641.

CASALI, J. G., and WIERWILLE, W. W., 1984, On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues. *Ergonomics*, **27**, 1033-1050.

COOPER, G. E., and HARPER, R. P., 1969, The use of pilot rating in the evaluation of aircraft handling qualities. AGARD Report 567, North Atlantic Treaty Organisation, Neuilly-sur-Seine, France.

CRABTREE, M. S., BATEMAN, R. P., and ACTON, W. H., 1984, Benefits of using objective and subjective workload measures. In *Proceedings of the Human Factors Society's 28th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 950-953.

DERRICK, W. L., 1981, The relationship between processing resource and subjective dimensions of operator workload. In *Proceedings of the Human Factors Society's 25th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 532-536.

DERRICK, W. L., and WICKENS, C. D., 1984, A multiple processing resource explanation of the subjective dimensions of operator workload. EPL-84-2/ONR-84-1, Engineering-Psychology Research Laboratory, University of Illinois, Urbana-Champaign, Illinois.

EGGEMEIER, F. T., 1981, Current issues in subjective assessment of workload. In *Proceedings of the Human Factors Society's 25th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 513-517.

EGGEMEIER, F. T., CRABTREE, M. S., ZINGG, J. J., REID, G. B., and SHINGLEDECKER, C. A., 1982, Subjective workload assessment in a memory update task. In *Proceedings of the Human Factors Society's 26th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 643-647.

EGGEMEIER, F. T., CRABTREE, M. S., and LAPOINTE, P. A., 1983, The effect of delayed report on subjective workload ratings. In *Proceedings of the Human Factors Society's 27th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 139-143.

EGGEMEIER, F. T., 1984, Workload metrics for system evaluation. In *Proceedings of the NATO Defence Research Group Panel VIII Workshop on 'Applications of Systems Ergonomics to Weapon System Development'* (Shrivenham, UK: Royal Military College of Science), pp. C5-C20.

EGGEMEIER, F. T., MELVILLE, B. E., and CRABTREE, M. S., 1984, The effect of intervening task

performance on subjective workload rating. In *Proceedings of the Human Factors Society's 28th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 954-958.

EGGEMEIER, F. T., and STADLER, M. A., 1984, Subjective workload assessment in a spatial memory task. In *Proceedings of the Human Factors Society's 28th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 680-684.

EGGLESTON, R. G., and QUINN, T. J., 1984, A preliminary evaluation of a projective workload assessment procedure. In *Proceedings of the Human Factors Society's 28th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 695-699.

EYSENCK, M. W., 1983, Anxiety and individual differences. In *Stress and Fatigue in Human Performance* (Edited by R. HOCKEY) (Chichester, UK: John Wiley & Sons Ltd.), pp. 273-298.

GALANTER, E., 1974, Psychological decision mechanisms and perception. In *Handbook of Perception: Volume 2- Psychophysical Judgement and Measurement* (Edited by E. C. CARTERETTE, and M. P. FRIEDMAN) (New York: Academic Press), pp. 85-125.

GARNER, W. R., 1974, Attention: The processing of multiple sources of information. In *Handbook of Perception, Volume II: Psychophysical Judgement and Measurement* (Edited by E. C. CARTERETTE, and M. P. FRIEDMAN) (New York: Academic Press), pp. 23-59.

GARTNER, W. B., and MURPHY, M. R., 1976, Pilot workload and fatigue: A critical survey of concepts and assessment techniques. NASA TN D-8365, National Aeronautics and Space Administration, Ames Research Center, Moffett Field, California.

GERATHEWOHL, S. J., 1975, Definition and measurement of perceptual and mental workload in aircrews and operators of air force weapon systems: A status report. In *Higher Mental Functioning in Operational Environments (AGARD-CPP-181)* (Neuilly-sur-Seine, France: North Atlantic Treaty Organisation), pp. C1-1 to C1-7.

GOPHER, D., and BRAUNE, R., 1984, On the psychophysics of workload: Why bother with subjective measures? *Human Factors*, **26**, 519-532.

HAIDER, M., 1972, Comparison of objective and subjective methods of the measurement of mental workload. In *Displays and Controls: The Proceedings for an Advanced Study Institute* (Edited by R. K. BERNOTAT, and K. P. GÄRTNER) (Amsterdam, The Netherlands: Swets & Zeitlinger N.V.), pp. 17-27.

HARRIS, C. S., and HABER, R. N., 1963, Selective attention and coding in visual perception. *Journal of Experimental Psychology*, **65**, 328-333.

HART, S. G., BATTISTE, V., and LESTER, P. T., 1984, Popcorn: A supervisory control simulation for workload and performance research. In *Proceedings of the Twentieth Annual Conference on Manual Control (NASA-CP-2341)* (Moffett Field, California: National Aeronautics and Space Administration), pp. 431-453.

HELM, W. R., 1981, Psychometric measures of task difficulty under varying levels of information load. In *Proceedings of the Human Factors Society's 25th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 518-521.

HENDY, K. C., 1979, Arousal and task performance: the ubiquitous U-curve. ARL-SYS-NOTE 68, Aeronautical Research Laboratories, Melbourne, Australia.

HENDY, K. C., 1986, A proposed programme of research: Systems Studies Group- Human Engineering Section. Technical Memorandum HF-86-10, Defence and Civil Institute of Environmental Medicine, Downsview, Ontario, Canada.

HICKS, T. G., and WIERWILLE, W. W., 1979, Comparison of five mental workload assessment procedures in a moving-base driving simulator. *Human Factors*, **21**, 129-143.

HOCKEY, R., and HAMILTON, P., 1983, The cognitive patterning of stress states. In *Stress and Fatigue in Human Performance* (Edited by R. HOCKEY) (Chichester, UK: John Wiley & Sons Ltd.), pp. 331-362.

HOLDING, D. H., 1983, Fatigue. In *Stress and Fatigue in Human Performance* (Edited by .R. HOCKEY) (Chichester, UK: John Wiley & Sons Ltd.), pp. 145-167.

JAMES, W., 1890, *The Principles of Psychology* (New York: Henry Holt and Co.).

JOHANNSEN, G., 1979, Workload and workload measurement. In *Mental Workload: Its Theory and Measurement* (Edited by N. MORAY) (New York: Plenum Press), pp. 3-11.

JOHANSSEN, G., MORAY, N. P., PEW, R., RASMUSSEN, J., SANDERS, A., and WICKENS, C.

D., 1979, Final report of the experimental psychology group. In *Mental Workload: Its Theory and Measurement* (Edited by N. P. MORAY) (New York: Plenum Press), pp. 101-114.

KAHNEMAN, D., 1973, *Attention and Effort* (Englewood Cliffs, New Jersey: Prentice-Hall).

KLAUSNER, S. Z., 1964, Concepts and indicators in studies of behavior under stress. AFOSR-640137, Air Force Office of Scientific Research, Washington D.C.

LEVISON, W. H., 1979, A model for mental workload in tasks requiring continuous information processing. In *Mental Workload: Its Theory and Measurement* (Edited by N. MORAY) (New York: Plenum Press), pp. 189-218.

MEISTER, D., 1985, *Behavioral Analysis and Measurement Methods* (New York: John Wiley and Sons).

MESHKATI, N., HANCOCK, P. A., and ROBERTSON, M. M., 1984, The measurement of human mental workload in dynamic organizational systems: An effective guide for job design. In *Proceedings of the First International Conference on 'Human Factors in Organizational Design and Management'*, pp. 521.

MORAY, N., 1967, Where is capacity limited? A survey and a model. *Acta Psychologica*, **27**, 84-92.

MORAY, N. (Ed.), 1979, *Mental Workload: Theory and Measurement* (New York: Plenum Press).

MORAY, N., 1982, Subjective mental workload. *Human Factors*, **24**, 25-40.

NAVON, D., and GOPHER, D., 1979, On the economy of the human-processing system. *Psychological Review*, **86**, 214-255.

NORMAN, D. A., and BOBROW, D. G., 1975, On data-limited and resource-limited processes. *Cognitive Psychology*, **7**, 44-64.

NOTESTINE, J. C., 1984, Subjective workload assessment and effect of delayed ratings in a probability monitoring task. In *Proceedings of the Human Factors Society's 28th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 685-689.

NYGREN, T. E., 1982, Conjoint measurement and conjoint scaling: A users guide. AFAMRL-TR-82-22, U.S. Air Force Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio.

O'DONNELL, R. D., and EGGEMEIER, F. T., 1982, Conceptual framework for the development of workload metrics in sustained operations. In *Sustained Intensive Air Operations: Physiological and Performance Aspects (AGARD CP-338)* (Neuilly-sur-Seine, France: North Atlantic Treaty Organisation), pp. 4-1 to 4-10.

OGDEN, G. D., LEVINE, J. M., and EISNER, E. J., 1979, Measurement of workload by secondary tasks. *Human Factors*, **21**, 529-548.

PARROT, J., 1971, The measurement of stress and strain. In *Measurement of Man at Work* (Edited by W. T. SINGLETON, J. G. FOX, and D. WHITFIELD) (London, UK: Taylor and Francis Ltd.), pp. 27-33.

PEPERMANS, R. G., and CORLETT, E. N., 1983, Cross-modality matching as a subjective assessment technique. *Applied Ergonomics*, **14**, 169-176.

RASMUSSEN, J., 1979, Reflections on the concept of operator workload. In *Mental Workload: Its Theory and Measurement* (Edited by N. MORAY) (New York: Plenum Press), pp. 29-40.

RASMUSSEN, J., 1983, Skills, rules and knowledge; signals, signs and symbols and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics*, **SMC-13**, 257-266.

REID, G. B., SHINGLEDECKER, C. A., NYGREN, T. E., and EGGEMEIER, F. T., 1981, Development of multidimensional subjective measures of workload. In *Proceedings of the IEEE International Conference on 'Cybernetics and Society'* (New York: IEEE), pp. 403-406.

REID, G. B., SHINGLEDECKER, C. A., and EGGEMEIER, F. T., 1981, Application of conjoint measurement to workload scale development. In *Proceedings of the Human Factors Society's 25th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 522-525.

REID, G. B., EGGEMEIER, F. T., and NYGREN, T. E., 1982, An individual differences approach to SWAT scale development. In *Proceedings of the Human Factors Society's 26th Annual Conference* (Santa Monica, California: Human Factors Society), pp. 639-642.

REID, G. B., SHINGLEDECKER, C. A., HOCKENBERGER, R. L., and QUINN, T. J., 1984, A projective

application of the subjective workload assessment technique. In *Proceedings of the IEEE National Aerospace and Electronics Conference (NAECON)* (New York: IEEE), pp. 824-826.

REVESMAN, M. E., and ROKICKI, S. M., 1985, Assessment of SWAT accuracy. In *Proceedings of the Human Factors Society's 29th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 183-187.

ROBINSON, C. P., and EBERTS, R. E., 1987, Comparison of speech and pictorial displays in a cockpit environment. *Human Factors*, **29**, 31-44.

ROHMERT, W., 1971, An international symposium on objective assessment of workload in air traffic control tasks. *Ergonomics*, **14**, 545-547.

ROLFE, J. M., 1971, The secondary task as a measure of mental load. In *Measurement of Man at Work* (Edited by W. T. SINGLETON, J. G. FOX, and D. WHITFIELD) (London, UK: Taylor and Francis Ltd.), pp. 135-148.

ROUSE, W. B., 1980, *Systems Engineering Models of Human-Machine Interaction* (New York: Elsevier North Holland).

SANDERS, A. F., 1979, Some remarks on mental load. In *Mental Workload: Its Theory and Measurement* (Edited by N. MORAY) (New York: Plenum Press), pp. 41-77.

SANDERS, A. F., 1983, Towards a model of stress and human performance. *Acta Psychologica*, **53**, 61-97.

SCHAFFER, L. H., 1971, Attention in transcription skill. *Quarterly Journal of Experimental Psychology*, **23**, 107-112.

SCHÖNPFLUG, W., 1983, Coping efficiency and situational demands. In *Stress and Fatigue in Human Performance* (Edited by R. HOCKEY) (Chichester, UK: John Wiley & Sons Ltd.), pp. 299-330.

SENDERS, J. W., 1983, *Visual Scanning Processes* (Tilburg, The Netherlands: PhD Thesis, Katholieke Hogeschool).

SHINGLEDECKER, C. A., CRABTREE, M. S., and ACTON, W. H., 1982, Standardized tests for the evaluation and classification of workload metrics. In *Proceedings of the Human Factor Society's 26th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 648-651.

SHINGLEDECKER, C. A., 1984, A task battery for applied human performance assessment research. AFAMRL-TR-84-071, U.S. Air Force Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio.

SINGLETON, W. T., 1971, The measurement of man at work with particular reference to arousal. In *Measurement of Man at Work* (Edited by W. T. SINGLETON, J. G. FOX, and D. WHITFIELD) (London, UK: Taylor and Francis Ltd.), pp. 17-25.

SKIPPER, J. H., RIEGER, C. A., and WIERWILLE, W. W., 1986, Evaluation of decision-tree rating scales for mental workload estimation. *Ergonomics*, **29**, 585-599.

STERNBERG, S., 1969, The discovery of processing stages: extensions of Donders' method. *Acta Psychologica*, **30**, 276-315.

STEVENS, S. S., 1962, The surprising simplicity of sensory metrics. *American Psychologist*, **17**, 29-39.

TAYLOR, M. M., 1986, Layered protocols for computer-human dialogue I: Principles. *International Journal of Man-Machine Studies*. In press.

THORNTON, D. C., 1985, An investigation of the Von Restorff phenomenon in post-test workload ratings. In *Proceedings of the Human Factors Society's 29th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 760-764.

VIDULICH, M. A., and TSANG, P. S., 1985, Assessing subjective workload assessment: A comparison of SWAT and the NASA-Bipolar methods. In *Proceedings of the Human Factors Society's 29th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 71-75.

WEINER, J. S., 1982, The measurement of human workload. *Ergonomics*, **25**, 953-965.

WICKENS, C. D., 1980, The structure of attentional resources. In *Attention and Performance* (Edited by R. S. NICKERSON) (Hillsdale, New Jersey: Laurence Erlbaum Associates), pp. 239-257.

WICKENS, C. D., 1984, *Engineering Psychology and Human Performance* (Columbus, Ohio:

Charles E. Merrill Publishing Company).

WIERWILLE, W. W., and WILLIGES, R. C., 1978, Survey and analysis of operator workload assessment techniques. S-78-101, Naval Air Test Center, Pataxent River, Maryland.

WIERWILLE, W. W., 1979, Physiological measures of aircrew mental workload. *Human Factors*, **21**, 575-593.

WIERWILLE, W. W., and CASALI, J. G., 1983, A validated rating scale for global mental workload measurement applications. In *Proceedings of the Human Factors Societies' 27th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 129-133.

WIERWILLE, W. W., and CONNOR, S. A., 1983, Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. *Human Factors*, **25**, 1-16.

WILLIGES, R. C., and WIERWILLE, W. W., 1979, Behavioral measures of aircrew mental workload. *Human Factors*, **21**, 549-574.

WINGERT, J. W., 1973, Function interlace modifications to analytic workload prediction. In *Crewstation Design: An Interagency Conference* (Edited by K. CROSS, and J. MCGRATH) (Santa Barbara, California: Anacapa Sciences), pp. 373-378.

YEH, Y. Y., and WICKENS, C. D., 1985, The effect of varying task difficulty on subjective workload. In *Proceedings of the Human Factors Society's 29th Annual Meeting* (Santa Monica, California: Human Factors Society), pp. 765-769.

Appendix 1: DEFINITION OF WORKLOAD RELATED TERMS

Attention

The concept of attention was described by James (1890), one of the first modern experimental psychologists, in the following terms: 'Focalisation, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others.' (p. 403).

Attention is the mental faculty that allocates resources to certain demands and ignores others. It is the means by which information is selectively perceived and processed by the operator. Years of research have been devoted to the understanding of attention, specifically with respect to the manner by which the operator is able to successfully attend some stimuli and filter out others. The stage of information processing (e.g., perception, sensory storage, or short-term storage) at which irrelevant or unwanted stimuli are filtered is not known.

Automation

Automation refers to unconscious cognitive processes that require minimal resources to maintain (Norman and Bobrow 1975, Sanders 1983). Automation occurs when an activity is fully learned; it corresponds to Rasmussen's term *skill-based behaviour* (Rasmussen 1983). Automation also implies a high order of stimulus response compatibility. In general, '...resources may be activated automatically when inputs to the system are highly compatible with outputs available to the resource.' (Hockey and Hamilton 1983, p. 356).

Capacity

The human operator is assumed to have an upper limit on the amount of cognitive resources that can be tapped, or on the amount of information that can be handled at any one time (Hockey and Hamilton 1983, Moray 1967). This limit is termed the operator's *capacity*. Under the multiple resources theory, a separate capacity exists for each distinct resource pool. Thus the operator actually possesses several capacities, delimiting each of his resources. Some authors presume that an overall attentional capacity, or effort, exists in addition to resource limits (Hockey and Hamilton 1983).

It is generally accepted that operator capacities vary both within and among operators, due to factors such as stress and fatigue (Eysenck 1983, Hockey and Hamilton 1983, Schönpflug 1983), and individual abilities. This renders the measurement of capacity an elusive pursuit.

Channel

Garner (1974) defines a channel as '...any property of a stimulus...that makes information available but is not itself part of the information aspect of the stimulus.' (p. 26). Rephrased, a channel is a conceptual medium through which it is assumed information is transmitted. The multiple resources theory postulates the existence of several parallel channels at any one stage of processing.

Code

The process by which sensory information is assigned to a single perceptual category, at which point the stimulus is said to have been perceived or recognized (Wickens 1984). There are various coding schemes that people may employ, e.g., spatial or verbal (Derrick and Wickens 1984), colour (Kahneman 1973), and object or dimension (Harris and Haber 1963). From a review of several studies of attention and memory, Kahneman (1973) concludes that '...subjects tend to develop the simplest possible codes that will suffice for the task at hand' (p.102).

Demand

Demand refers to the amount of resources required to achieve a certain level of performance under given subject-task parameters (Navon and Gopher 1979). Demands can be either external or internal in origin, although as Schönpflug (1983) points out, '...external demands...cannot operate within an individual unless they have been identified by him and internalised to become part of his set of internal demands.' (p. 301). Task demand is directly related to the amount of resources required and the level of intended performance (Navon and Gopher 1979). A particularly notable shortcoming of the measurement of human performance and the evaluation of man-machine systems, is the researcher's inability to properly assess task demands in either an absolute or a relative sense.

Effort

Effort refers to the intensity of processing, or the amount of cognitive resources expended by the operator in the performance of a task. Effort is viewed in this report as being entirely under the volition of the operator. Although an operator may exert less effort as he becomes cognitively fatigued (Holding 1983), this is believed to be a voluntary diversion of attention from the task. Similarly, operator effort may be increased to compensate for the reduced capacity effects that accompany fatigue (Schönpflug 1983), in order to maintain performance at an acceptable (perhaps constant) level.

There is agreement on the proposition that effort and attention are closely related (e.g., Eysenck 1983, Hockey and Hamilton 1983), however the extent of this relationship is controversial. Hockey and Hamilton describe effort as being the *executive resource*. By this, they are referring to the attentional resource that allocates all other attentional resources to various demands.

Mental Workload

The term workload can refer to cognitive (or mental) load, physical load, or both (sometimes termed *human operator load*). The focus of this report is on mental load, and subjective mental load in particular. Thus any reference to workload in this report, unless explicitly stated otherwise, is concerned primarily with mental load. The extent to which physical load cannot be distinguished from mental load is a complicating factor in workload measurement. The utility of segregating the two phenomena will not be addressed here.

Modality

Modality refers to the medium of communication between the operator and the outside world. Examples of input modalities are visual, auditory, kinaesthetic, olfactory, gustatory, and tactile. The dominant output modalities are vocal and manual responses although other motor responses are possible.

Motivation

The term *motivation* refers to the operator's willingness to exert effort in the performance of a task. Galanter (1974) offers the following remarks about motivation: 'On the one hand, motivation may refer to purely psychological desires, aversions, feelings, attitudes, interests...On the other hand...the concept of specific motivations is objectified by construing them as the consequences of presumably objectively determinable changes in the physiological state of the organism' (pp. 85-86). To simplify the latter portion of this claim, any traceable physiological change is believed to be accompanied by specific motivations related to that change. The example offered by Galanter is the following: denial of food to the human organism causes hunger, which is accompanied by the motivation to eat.

Resources are dispensed among various demands according to the operator's motivations, i.e. according to subjective feelings of the value of performing various functions. These feelings are based on previous experiences, which have caused mental models to be developed and stored in the operator's long-term memory. There is a direct relationship between motivation, effort, attention, and resources: the higher one's motivation is to perform a task, the greater the effort expended towards the performance of that task, more attention is paid to that task, and more resources are dedicated to the task. Assuming that the task is performed with a constant strategy over these ranges of motivation, effort, attention, and resources, performance will increase along with these parameters up to the workload threshold level. In the overload situation, performance may decrease with increased demand, according to the Yerkes-Dodson Law in arousal theory (Schönpflug 1983). The Yerkes-Dodson Law postulates an inverted U-shaped function relating performance to arousal (activation) or demands (Hockey and Hamilton 1983). Note however, that the treatment of arousal in this way, as a simple uni-dimensional construct, is open to question (Hendy 1979). Along with this, it is probable that effort will decrease in an overload situation (Schönpflug 1983).

Overload

When the imposed task demands require processing resources which exceed the operator's capacity, *overload* is said to occur.

Response Output

Any observable response by the operator, to an external stimulus or demand, is termed the ***response output***. Outputs can be both voluntary and involuntary.

Resources

The concept of *resources*, as described by Norman and Bobrow (1975) and Navon and Gopher (1979), will be central to the discussion and development of a model for mental workload. It is assumed that the human operator draws upon processing pools or reservoirs, called resources, in the cognitive performance of a task. These resources are, by nature, limited both in number and in capacity. The allocation of an attentional resource is assumed to be under some type of cognitive control (Wickens 1980) although the resources themselves, and their allocation, are not part of our awareness (Derrick and Wickens 1984). As suggested by Allport, Antonis and Reynolds (1972), Wickens (1980), and Navon and Gopher (1979), these resources are multiple; certain resource pools are called upon to perform certain functions, leaving others untapped until they are required to perform the function(s) to which they are best suited. For example, resources used for manual control are called on when the human operator is engaged in a manual tracking task; resources appropriate for memorising a series of digits are not employed to perform the tracking task and therefore are, theoretically, available for a concurrent memory task.

The actual designation of resources to functions or behaviours is not yet well understood, although there have been some assertions. Notable is the resource classification scheme of Wickens (1980). Wickens proposed three independent and dichotomous dimensions, hence six separate pools of resource, *viz.*: (1) processing stages (consisting of the dichotomy of ***perceptual evaluation and central processing*** on one hand and ***selection and execution of response*** on the other); (2) processing modality (***visual***, or ***auditory*** input); and (3) processing codes (***vocal/verbal***, or ***manual/spatial*** guided responses). For example, assuming resources are allocated according to the structure of ***processing stages***, one would not expect that the performance on a tracking task which requires selection of response resources, would be unduly disrupted by the concurrent peformance of a mental arithmetic task which requires central-processing resources in working memory (Wickens 1984).

Multiple resource theory draws on *resource volume* models of human behaviour. It assumes process invariance (Sanders 1983), i.e. the operator may alter the allocation of resources to demands (e.g., in a dual task environment) but the manner in which the tasks are performed does not change. The alternative to this, *resource strategy*, argues that varying demands not only alter the allocation of resources, but also the fundamental operations involved in performing a task.

Stages of Processing

The term *stages of processing* (alternatively *processing stages*) refers to conceptual components of information processing that are hypothesised to be discrete in nature (Sternberg 1969). The basis of the componential divisions may be temporal, functional, or both. The processes of *perception and central processing* (translation, memory storage and retrieval, and decision making) on one hand, and *response selection and execution* on the other, have been identified by numerous studies as candidates for discrete stages (e.g., Derrick 1981, Schaffer 1971, Wickens 1980).

Stress and Strain

A number of authors (e.g., Parrot 1971, Singleton 1971, Weiner 1982) have drawn a definite distinction between the two often confused terms *stress* and *strain*. These authors suggest that stress is the operator loading aspect of the man-machine system. Although Sanders (1983) considers this definition to lack '...common sense meaning' (p. 62), it is valuable as a working definition in performance theory. The cause of the stress, or stimulus, is termed the *stressor* (Sanders 1983). Weiner (1982) takes the position that 'Strain has been defined as constituting those changes indicative of homeostatic disturbance which appear when the stress responses become insufficient to cope with the imposed load.' (p. 961). A helpful analogy can be drawn from the mechanics of solids. If a load is applied to an object as a stress, then a deformation (strain) will result. Once this load exceeds certain limits, the deformation which takes place cannot be compensated for by the material's elastic properties, and the object's dynamic and internal behaviour is altered. Strain cannot be present without stress as the direct cause. Just as the inanimate object's properties are changed when strained beyond the elastic limit, the human operator who is cognitively strained is affected in mental behaviour.

Parrot (1971) qualifies his definitions by stating that strain is a necessary effect if a load is considered to constitute a stress. Weiner (1982) draws on the mechanical analog of plastic and elastic deformation to make the distinction between stress and strain. It is Weiner's position which is adopted in this paper, i.e. stress exists as an external force acting on the operator; if the stress is excessive and homeostatic limits are exceeded, a strain results. In this context the existence of a phsyiological/pyschological strain is equated with non-elastic deformation, at least in the short term.

Task Inputs

Task inputs refer to those demands placed on the operator by the task, which are perceived and internalized.

Threshold

The performance threshold is the point at which the task demands are equal to the operator's capacity to perform that task, i.e. the point at which the operator is fully loaded. Below threshold level it is assumed that the operator can cope with the task demands; above the threshold, the operator's performance is degraded (Norman and Bobrow 1975).